

修士論文

異常値検出を用いた
文化差検出の改善

指導教官 村上 陽平 准教授

立命館大学大学院情報理工学研究科
計算機科学コース

西村 一球

2022年1月31日

異常値検出を用いた文化差検出の改善

西村 一球

内容梗概

近年、グローバル化によって異なる言語・文化を持つ人との会話が増えつつある。機械翻訳の品質向上により多言語コミュニケーションの言語の壁は解消されつつあるが、依然として文化差による誤解が生じている。例えば、日本でよく食べられる“ゴボウ”があげられる。“ゴボウ”を機械翻訳で英語に翻訳すると“burdock”という結果が得られる。日本人の多くは“ゴボウ”と聞くと根っこの見た目の食材を想起する。一方で、海外では“burdock”からは赤く特徴的な花をつける植物を想起することが多い。想起されるイメージは全く異なるものであるが、これらは同一の植物である。“ゴボウ”とは“burdock”の根の部分であり、日本人はこれを食べる食文化があるため“ゴボウ”は根っこの見た目の食べ物であると認識している。一方で、海外では“ゴボウ”を食べる文化が存在しないため、根っこの部分ではなくより特徴的な花の部分想起する。

そこで、我々は一つの概念に紐づけられた日英それぞれの単語から画像検索を行い、取得された画像の類似性に基づいて文化差検出を行う。ただし、取得した画像にターゲットと異なる画像が含まれる可能性があるため、異常値検出手法を用いて各画像の確信度を算出し、確信度に基づくノイズの除去と画像ベクトルの生成を行った。また、本手法を正確に評価するために、日英の母語話者に各画像集合で表されるイメージの想起の有無をラベル付けしてもらい、母語話者間で比較することで、文化差の有無を判定し、文化差の評価データセットを作成する。本手法を実現するために、本研究で取り組んだ課題は以下の2点である。

文化差判定用の評価データセットの構築

日本語から取得した画像と英語の単語で取得した画像をもとに日本語話者のみが文化差判定を行うと、英語話者の視点や言葉といった要素が含まれていない。また、文化差とは形の違いだけではなく、用途の違いなども含まれる。従って、画像だけではなく単語や説明といった複数の要素をもとに日英の両方の視点から対象概念の文化差判定を行なった評価データセットを作成する必要がある。

ノイズデータに頑健な文化差検出手法

画像をもとに文化差の有無を判定するため、画像にターゲットと異なるノ

イズ画像が含まれている場合、判定を間違える可能性がある。そのため、ノイズとなる画像に影響されない文化差検出手法を作成する必要がある。

一つ目の課題である評価データセットの作成のために、概念辞書から概念を1000個取り出し、各概念ごとに日英の単語群、概念の説明、単語群から画像検索によって取得された画像を日英の言語話者に見せる。各話者は母言語の情報をもとに取得された画像が単語や説明からイメージできるかを5段階で評価する。言語によってこれらの概念の評価に差が見られる概念を文化差あり、差が見られない概念を文化差なしとしてラベリングする。二つ目の課題であるノイズデータに頑健な手法では、取得された画像から特徴ベクトルを生成し、それらをクラスタリングによって分類し、分類されたクラスターの要素数に応じて特徴ベクトルに重み付けを行う。これをもとに各言語の合成ベクトルを生成し、生成されたベクトル間の類似度によって文化差判定することでノイズデータに影響されにくい文化差検出をする。本研究の貢献は以下の通りである。

文化差判定用の評価データセットの構築

日英の両言語話者の視点から言葉と画像の複数の要素を評価をもとに差が見られる概念を文化差あり、差が見られない概念を文化差なしとしてラベル付けした。日英間で文化差ありが257件、文化差なしが565件獲得できた。また、1000個の概念を概念辞書の上位概念に基づいて分類した結果、分類したクラスごとに獲得できた文化差ありの個数は、object:16, matter:9, process:27, causal agent:22, thing:25, psychological feature:31, attribute:10, group:23, communication:40, measure:23, relation:31であった。

ノイズデータに頑健な文化差検出手法

作成した評価データセットを用いて提案手法を評価した結果、ノイズデータの除去や画像の確信度を用いない手法と比較して、Accuracyは向上しなかったものの、これにより文化差判定結果が良くなるケースも見受けられた。

Improving Cultural Difference Detection using Outlier Detection

Ikkyu Nishimura

Abstract

These days, the problems caused by language difference in multilingual communication have been resolved by the improvement of machine translation, but there are still misunderstandings due to cultural differences. For example, “ゴボウ (gobou)” which is often eaten in Japan is translated by machine translation as “burdock”. People in many countries besides Japan usually associate “burdock” with an inedible plant above the ground. The two images triggered by these words are all correct. Because both are the same plant.

Previously, cultural difference detection method based on similarity of images on web has been proposed. However, the acquired images may contain different images from the target. In this research, we calculated the confidence of each image using the outlier detection method, by removing noise data and generating the image vector. In order to accurately evaluate this method, we asked native speakers of Japanese and English to label the presence or absence of recall of the images represented by each image set. By comparing the difference between the results from native speakers of each language, we created an evaluation dataset that determines the presence or absence of cultural differences. We address the following two issues in this research.

Construction of evaluation dataset for cultural difference detection

If only Japanese speakers make a cultural difference judgment based on the images acquired in each language, elements such as the viewpoint of English speakers and English words are not included. In addition, cultural differences include not only the differences in shape but also differences in usage. It is necessary to create an evaluation dataset that judges the cultural difference of the target concept from both Japanese and English perspectives based on multiple elements such as words and explanations as well as images.

A robust method for cultural differences detection with noisy data

In order to judge whether a cultural differences exists or not based on the images, if the images include some noise images, the images that are different from the target, the judgment could be wrong. Therefore, it is necessary to create a cultural difference detection method that would not be affected by the noisy images.

To create the evaluation dataset, we took 1000 concepts from the concept dictionary and displayed Japanese-English words, explanations, and images obtained from the words to Japanese and English language speakers for each concept. Each language speaker evaluated whether the image acquired based on the information in the mother language can be imagined from words and explanations on a five-point scale. We labeled the concepts that show differences in the evaluation of concepts depending on the language as cultural differences, and the concepts that do not show differences as no cultural differences. In this robust method for noisy data, we generated feature vectors from the acquired images and classified them using a clustering method. We weighted the feature vectors according to the number of elements in the classified clusters. We used this to generate a composite vector for each language, and by determining the cultural difference based on the similarity between the vectors, it is possible to detect cultural differences that are not easily affected by noise data. The contributions of this research are to construct the dataset evaluation and to improve the accuracy of the cultural difference detection.

Construction of evaluation dataset for cultural difference detection

We labeled concepts that differed in the evaluation of concepts from the perspectives of both Japanese and English speakers as having cultural differences, and concepts that did not differ as having no cultural differences. We were able to obtain an evaluation data set that included 257 cases with cultural differences and 565 cases with no cultural differences.

Robust method of cultural differences detection for noisy data

As a result of evaluating the proposed method using the created evaluation dataset, the accuracy did not improve compared to the method that did not use noise data removal or image certainty, but this improved the cultural difference judgment result. Cases were also seen.

異常値検出を用いた文化差検出の改善

目次

第1章	はじめに	1
第2章	多言語コミュニケーションにおける文化差	3
2.1	文化差とは	3
2.2	文化差の体系	4
2.2.1	文化差あり(排他関係)	5
2.2.2	文化差あり(部分的共通関係)	5
2.2.3	文化差あり(包含関係)	5
2.3	関連研究	6
2.3.1	知識に基づく文化差研究	6
2.3.2	イメージに基づく文化差研究	7
第3章	画像特徴量に基づく文化差検出手法	8
3.1	特徴ベクトルに基づく類似度算出	8
3.2	閾値を用いた文化差判定	8
3.3	予備評価	10
3.3.1	評価手順	10
3.3.2	最適値の検証	11
3.4	予備評価の分析	13
3.5	既知の文化差への適用	15
第4章	評価データセット	18
4.1	評価データセットの作成	18
4.2	アンケート	20
4.3	評価データセットの集計	21
第5章	異常値検出を用いた改善	26
5.1	K-means	26
5.2	DBSCAN	27
5.3	IsolationForest	28
第6章	評価実験	29

6.1	実験の手順	29
6.2	実験結果	29
第7章	考察	38
7.1	評価データセットの分析	38
7.2	文化差検出結果の例	39
7.3	従来手法と改善手法の比較の例	46
7.4	今後の改善	48
第8章	おわりに	50
	参考文献	52
8.1	付録	54

第1章 はじめに

2015年国連によって定められたSDGs¹⁾(Sustainable Developments Goals)では、持続可能でよりよい世界を目指すをコンセプトに17のゴール、169のターゲットの開発目標が記されている。近年、国際的な諸問題に対して社会の多様性を考慮しながら取り組むために、言語や文化の違いを超えた異文化コラボレーションが求められている。こうした能力を育む地球市民教育の実現がSDGsの一つとして規定されている。例えば、NPO Pangaeaは地球市民教育の実現をテーマに、異なる言語・文化を持つ児童を集めて、協働で世界の諸問題の解決策をデザインする異文化サマースクール(KISSY:Kyoto Intercultural Summer Schools for Youths)²⁾を開催している。

このような異文化コラボレーションでは、近年、機械翻訳の翻訳精度改善により言語の差は解消されつつあるが、依然として文化差による誤解が生じている。例えば、KISSYでは独自の機械翻訳ツールを用いて会話が行われている。その中で日本語話者の児童が“あんこ”を翻訳した際に、得られた英語の翻訳結果が“red beans paste”であった。これを用いてクメール語話者の児童にあんこを説明したときクメール語話者が想起するイメージが異なった。日本人が想起するペースト状の“あんこ”に対して、“red beans paste”からはより四角く固められたお菓子のようないあんこが想起され、誤解を生じさせることがあった[1, 2]。そこで、このような文化差による問題を解決するために、私たちはWeb上の画像を用いた文化差検出手法を提案している。この手法では、多義語の場合、その単語の訳語と異なる語義の画像が収集され文化差の検出を誤る可能性があるため、訳語と同じ語義を表す画像を収集し、その画像間の特徴の差に基づいて文化差検出を行う。また、画像特徴量をもとに文化差検出するには、言語間で画像特徴量の類似度を計算し、その類似度により文化差の有無を判定する。しかしながら、現状の問題として画像検索によってターゲットと異なる画像が取得された場合、それらの影響により正確な文化差判定ができない可能性がある。そのため、本研究では、異常値検出を用いてノイズデータに頑健な文化差検出手法を提案している。具体的には、同一概念に紐づけられる日英それぞれの単語から画像検索を行い、取得された画像の特徴ベクトルをクラスタリングによってノイズ検出を

¹⁾ <https://sustainabledevelopment.un.org/?menu=1300>

²⁾ https://www.pangaeaan.org/web/japanese/general/aboutpangaeact_jp.html

行う。その結果をもとに特徴ベクトルに重みを付与し、各言語ごとに重み付きベクトルを加重平均によって合成する。生成されたベクトル間の類似度に基づいて文化差判定を行う。本手法を実現するために、本研究で取り組んだ課題は以下の2点である。

概念間の文化差を判定した評価データセット

日本語から取得した画像と英語の単語で取得した画像をもとに日本語話者のみが文化差判定を行うと、英語話者の視点や言葉といった要素が含まれていない。また、文化差とは形の違いだけではなく、用途の違いなども含まれる。従って、画像だけではなく単語や説明といった複数の要素をもとに日英の両方の視点から対象概念の文化差判定を行なった評価データセットを作成する必要がある。

ノイズデータに頑健な文化差検出手法

画像をもとに文化差の有無を判定するため、画像にターゲットと異なるノイズ画像が含まれている場合、判定を間違える可能性がある。そのため、ノイズとなる画像に影響されない文化差検出手法を作成する必要がある。

以下本項では、2章において、多言語コミュニケーションで生じる文化差を説明し、既存の文化差へのアプローチを紹介する。続いて、3章において、画像特徴量に基づく文化差検出手法を説明し、それを用いた予備評価の説明を行う。続いて、4章では概念間の文化差を判定した評価データセットの作成方法を説明し、5章で異常値検出を用いた改善手法を述べる。6章では提案手法の評価実験の実験手順の説明と評価を行い、その結果をもとに7章で提案手法の分析・考察を行う。

第2章 多言語コミュニケーションにおける文化差

2.1 文化差とは

ゴボウ



burdock



図1: 文化差の例

機械翻訳の精度向上によって、多言語コミュニケーションが可能になりつつあるが、いまだにコミュニケーションに齟齬が生じる場合がある。その理由の一つが文化差である。人は生まれ育った環境・触れてきた文化によって、人それぞれの文化的背景を持っている。それに伴い言葉やものからイメージするもの、考え方が異なってくる。これにより、会話において話し手と聞き手が想起しているイメージが異なりコミュニケーションに齟齬が生じる [3]。同じ国や地方であってもこうした文化差によるコミュニケーションの齟齬は起きることがあるが、異なる国・言語・文化の人と会話を行う場合、こうした文化的背景の差はより大きいものとなる。そのため、本研究では異なる国・文化・言語の人を対象に文化的背景の違いを文化差として検出することを目指す。

具体的な文化差の例として、日本でよく食べられる“ゴボウ”があげられる。機械翻訳で“ゴボウ”を翻訳すると“burdock”という翻訳結果が得られる。日本人の多くは“ゴボウ”と聞くと根っこのような見た目（図1の左）の食べ物を想起する。しかし、日本以外の多くの国では“burdock”からはイガのある特徴的な赤い花をつける植物（図1の右）を想起することが多い。これらの両方のワード



図2: 文化差なし

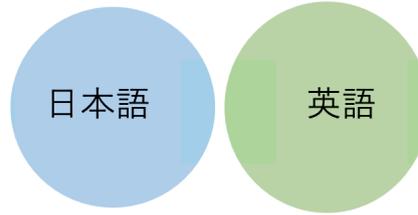


図3: 文化差あり (排他関係)

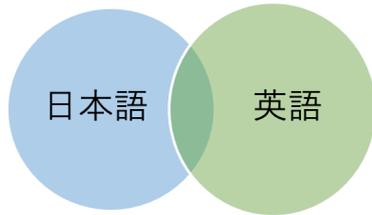


図4: 文化差あり (部分的共通関係)

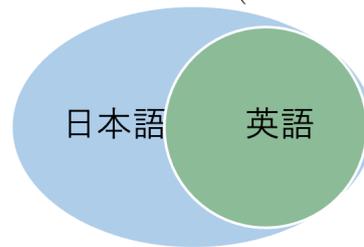


図5: 文化差あり (包含関係)

から想起されたイメージは大きく違っているが、どちらのイメージも間違っているわけではない。“ゴボウ”と“burdock”で得られるイメージのものは同じ植物であり、日本で言う“ゴボウ”は“burdock”の根っこの部分を指している。日本ではこの植物の根っこの部分を食べる文化を持っているため、このような根っこ(図1の左)のような見た目を想起し、食材と認識されている。一方で、海外では“ゴボウ”を食べる文化があまり存在しないため、根っこに比べてより特徴的な花の部分のイメージが想起される。

2.2 文化差の体系

文化差には様々なケースが存在し、文化の違いの程度の大きさが様々である。それらの程度の大小に応じて多言語コミュニケーション上で生じる問題が異なる。そのため、概念の範囲を用いて文化差を大きく4つのパターンに分け、それぞれに生じる問題と文化差の例を紹介する。まず、各言語の単語から想起されるイメージを円で範囲的に表すと図2, 3, 4, 5のように、日本語が表す概念範囲の円と英語が表す概念範囲の円が出来上がる。このとき、図2のように日英の概念範囲が一致している場合において、両言語話者が想起するものが同じとして文化差なしと言える。一方で、これらの言語ごとの概念範囲が異なっているときには文化差があると言える。次に文化差が存在するケースにおいて、3つのパターンを紹介する。

2.2.1 文化差あり (排他関係)

図3のように概念の範囲が完全に一致していない場合、このような文化差を排他関係とする。このような文化差の例として、“ゴボウ”が挙げられる。日本人が想起している“ゴボウ(図1左)”に対して、アメリカ人が想起している“burdock(図1右)”は完全に想起しているものが異なっている。これらは日本独自である“ゴボウ”を食べる文化が存在するために引き起こされる。このような排他関係の文化差の特徴として、多言語コミュニケーションにおける危険度は他のパターンに比べて低いと考えられる。なぜなら、排他関係の文化差は概念範囲が違いすぎるため、多言語コミュニケーション上で会話に齟齬が生まれる可能性が高い。一方で、あまりにも会話が成り立たなくなるため話者同士が誤解に気づく可能性も高いと言える。そのため、排他関係は多言語コミュニケーションにおける危険度は低い。

2.2.2 文化差あり (部分的共通関係)

図4のように概念範囲が部分的に一致している場合、このような文化差を部分的共通関係とする。この部分的共通関係は排他関係と次に説明する包含関係の中間的な位置付けである。このような文化差の例として、“餃子”が挙げられる。発祥の中国では餃子とは日本でいう水餃子として食べられる。一方で、日本では水餃子もあるが、餃子といえば焼き餃子のことを指している。このように日本独自の文化もあれば、中国では餃子は基本的には祝い事の席で食べられるという側面を持っている。このように部分的共通関係の文化差はある部分はお互いの文化に共通することもあるが、各文化によって独自の部分を持っている。部分的共通文化差は部分的には概念範囲が一致しているため、会話において話者同士が気づかずに誤解が生まれる可能性が高い。

2.2.3 文化差あり (包含関係)

図5のようにある一方の概念範囲がもう一方の概念範囲に包含されている場合、このような文化差を包含関係とする。このような文化差の例として、“クジラ”が挙げられる。日本人もアメリカ人も“クジラ”について想起しているものは同じである。しかしながら、日本人にはクジラを食べる食文化を持ち合わせているため、日本語で表す概念範囲が英語での概念範囲より少しだけ大きい。排他関係の例であげた“ゴボウ”との大きな違いは想起しているものが“クジラ”ではどちらの文化・言語でも想起されるイメージは同じである。このような包含関係の文化差はどちらか一方の概念範囲が大きいいため、どちらか一方には話が

通じているが、もう一方には齟齬が生まれている可能性がある。“クジラ”の例の場合、日本語話者にとっては“クジラ”が食用として話は通じているが、英語話者にとっては“クジラ”が食材とはつながらないため、会話に齟齬が生まれている。そのため、日本語話者にとっては文化差が存在することがわかりづらく、文化差の発見がしづらい。このように、どちらか一方からしか文化差があることが気付きづらいため、包含関係はコミュニケーション上での危険度が高い。

2.3 関連研究

多言語コミュニケーションの文化差に関する既存の研究は、知識に基づく文化差と、イメージに基づく文化差の2種類に大きく分けられる。

2.3.1 知識に基づく文化差研究

知識に基づく文化差研究とは、人の知識を体系化したものや人の体験をもとに文化差を明らかにする研究である。

吉野らは、文化差理解を支援するために、Web上にある人の体験談をもとに作られた「文化差の理解を促す事例」を文化差データとして収集する手法を提案している [4]。また、他にも吉野らは Wikipedia を用いた文化差検出手法を提案している [5]。この研究では、事前アンケートによって文化差が存在すると人手で判断された言葉に対して、Wikipedia 上のカテゴリなどの情報を用いて文化差を検出できるかを確認するものである。さらに、語句の重要度を計算することで文化差の検出精度が向上することが示されている [6]。これらは、Wikipedia の文章から対象語句を抽出することで文化差を検出するもので、文字レベルでの文化差検出手法である。

同じく、Wikipedia を利用した文化差に関する研究がある。Ulrike らは、フランス、ドイツ、日本、オランダの文化的多様性を Wikipedia に適用し、それぞれの Wikipedia の編集操作から国の文化とコンピューターを介したコミュニケーションの関係を調査している [7]。研究の結果、Hofstede が明らかにした文化的影響の4つの次元と相関していることがわかった [8]。

高崎らによる研究で他言語の話者に知識を伝える際のコミュニケーションに着目したものがある。この研究では、ベトナムの児童を介して日本の稲作の専門家からベトナムの農業事業者に農業の知識を伝えるプロジェクトをもとにしている。知識伝達の際に、仲介者を挟むことにより、専門家が意図していないニュアンスでベトナムの事業者に伝達される問題があったことを述べている [9]。

本研究では想起されるイメージの違いなど文字レベルではわからない文化差を考慮して、文化差判定のために画像の特徴ベクトルに基づく文化差検出手法を提案している。

2.3.2 イメージに基づく文化差研究

イメージに基づく文化差研究とは、画像や絵文字など視覚的に得られる情報をもとに文化差を明らかにしている研究である。

柳井らは写真共有サイト上のタグ付き画像を利用する文化差検出手法を提案している。概念の代表的な写真をタグの地域ごとに分類している。これによって、同一概念の対象物が地域ごとにどのような文化差があるのかを視覚的に明らかにしている [10]。

Pituxcoosuvarn らによる研究では、日本語と英語の単語をもとに画像検索を行い、それらの画像の特徴ベクトル間の類似度を算出する。算出された類似度をもとに人の主観で文化差を判定している [2]。この研究では本研究と同様に画像の特徴ベクトル間の類似度によって文化差を判定しているが、算出された類似度は連続値であるため、最終的には人手によって文化差を判定している。本研究では算出された類似度をもとに閾値を使って文化差を自動的に判定するとともに、ノイズとなる画像の影響をあまり受けずに文化差判定する手法を提案している。

Cho らは絵文字を用いて文化差の検出を試みている [11]。この研究はネットワークを介して異文化の話者が会話するときに言語を使わない意思伝達のツールとして絵文字に着目したものである。アメリカと日本の人にアンケートで120種類の絵文字それぞれの印象について調査したところ、絵文字は文化によって解釈のされ方が違うことが明らかにされている [12]。

神田らは近年のネットワークコミュニケーションでよく利用されるアバタに着目して、アバタの表情の解釈において文化差が存在するかを分析している。アジアと欧米8カ国間のアバタの表情の解釈内容を比較して分析したところ否定的な内容の解釈には文化差がなかったが、肯定的な内容の解釈には文化差が存在することを発見している [13]。

これらの研究は本研究と同様にイメージをもとに文化差を判定する研究であるが、判定自体は人手で行っている。一方で、本研究は文化差を画像特徴量を用いて自動的に検出する手法を提案する。

第3章 画像特徴量に基づく文化差検出手法

本研究では, 文化差を検出するために Web 上の画像の類似性に基づいて文化差判定を行なっている. 本章では, その文化差検出手法を具体的に説明する.

3.1 特徴ベクトルに基づく類似度算出

文化差を検出するために, WordNet³⁾と日本語 WordNet⁴⁾という概念辞書を用いる. WordNet では概念をもとにその概念を表すワードと説明, その概念の上位概念や下位概念が対応づけられていて, それらを日本人が人手で日本語対応させたものが日本語 WordNet である [14][15].

WordNet は各概念を同義語の集合である Synset で定義するため, 一つの概念に複数の単語が割り当てられる. 表 1 の日本語 WordNet の場合, “ゴボウ, 牛蒡”と “burdock, clotbur” がゴボウの概念を表す日英の単語である. これらの取得された単語群をキーワードにして, 画像検索を行い, 各概念につき上位 10 件の画像を取得する. ゴボウのように Synset が複数の単語で構成される場合は AND 検索を行うことで対象概念を表す画像群を収集する.

次に, Keras の VGG16³⁾を使って各画像の特徴ベクトルを生成する [16]. この時, $4 \times 4 \times 512$ 次元から 1 行 8192 次元のベクトルに整形する. 各言語ごとに画像 10 枚から生成された特徴ベクトルを合成するために, 単純平均を用いて特徴ベクトルの平均ベクトルを算出し, 言語ごとの対象概念の特徴ベクトルとする. その後, 平均特徴ベクトル同士でベクトル間の類似度を COS 類似度で計算する. 得られた類似度と閾値を比較して文化差の有無を判定する. 類似度が閾値を下回った場合は文化差ありと判定し, 閾値を上回った場合は文化差なしと判定する.

3.2 閾値を用いた文化差判定

閾値とは算出した類似度から文化差の有無を判定するための基準となる値のことである. 類似度はあくまで連続値であるために文化差を判定するための基準となる閾値を適切に設定する必要がある.

³⁾ <https://wordnet.princeton.edu>

⁴⁾ <http://compling.hss.ntu.edu.sg/wnja/>

³⁾ <https://arxiv.org/abs/1409.1556>

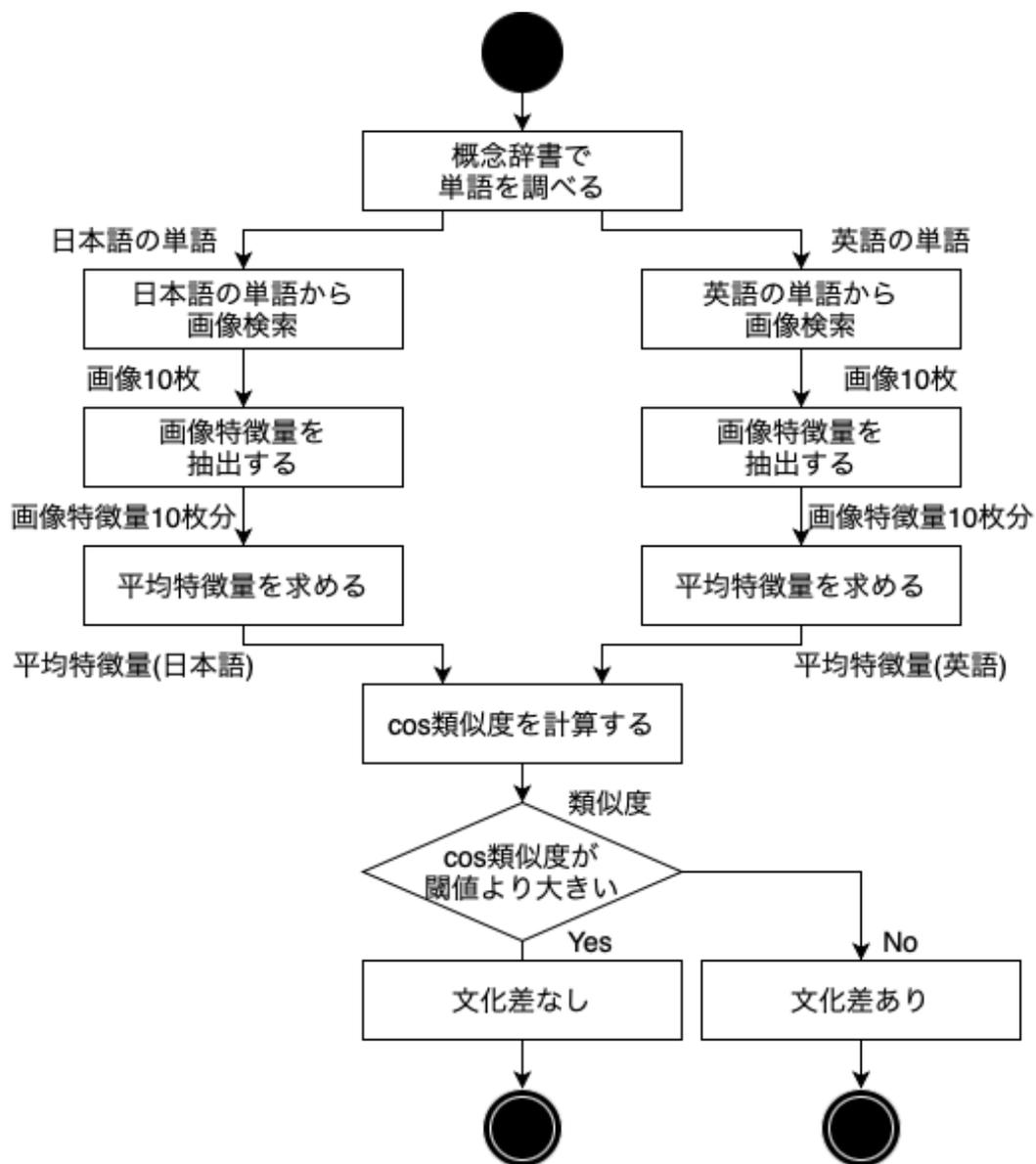


図 6: 文化差検出手法

まず、文化差の判定基準の閾値を最適化するために、文化差の検出精度の指標を定義する。検出精度の指標には文化差判定の正確さ (Accuracy) を用いる。Accuracy とは、提案手法の文化差判定結果が人手による文化差判定結果と一致した割合である。表 2 で True と表記している箇所が人手での判定と提案手法の判定が一致するパターンである。これをもとに提案手法の Accuracy の算式を以下に示す。

$$Accuracy = \frac{\text{True に該当した synset の数}}{\text{True と False に該当した synset の数}} \quad (1)$$

表 1: 日本語 WordNet の例

言語	Synset	説明
日本語	ゴボウ, 牛蒡	ユーラシアの温帯地域の直立性の2年生草本の総称で, しっかりした主根を持ち, いがをつける
英語	burdock, clotbur	any of several erect biennial herbs of temperate Eurasia having stout taproots and producing burs

表 2: 判定結果の比較

	文化差有り (人手)	文化差無し (人手)
文化差有り (提案手法)	True	False
文化差無し (提案手法)	False	True

これらの指標をもとに閾値を最適化するための予備評価を行う。具体的には、人手により文化差があると判定された概念 250 個と文化差がないと判定された概念 250 個の計 500 個の概念を用いて提案手法で文化差判定を行う。この時、閾値を 0.0 から 1.0 まで 0.05 ごとに変化させた時に、最も Accuracy が最大化する閾値を文化差検出に最適な閾値とする。

3.3 予備評価

3.3.1 評価手順

人手で文化差を判定した場合、人によって文化差の感じ方が異なるため、予備評価では一つの概念につきアンケートにおいて、計 5 人の日本人で文化差の有無を判定し、多数決によって一つの判定結果に定める。

図 38 は人手で文化差を判断するときに用いたアンケートである。アンケートの質問は次のとおりである。

キーワードを見て、“A” と “B” のどちらの画像群がより連想できるか選んでください。もし、両方とも連想できた場合は “両方” を選び、どちらも連想できない場合は “該当なし” を選んでください。※ () 内の数字は質問番号です

アンケートでは、WordNet からランダムに選択された 1000 個の Synset を用いた。アンケートでは、キーワードと画像群 A と画像群 B の三つの情報を回答者に与えている。図 38 のようにアンケートの選択肢には、“A”、“B”、“両方”、“該当なし” の 4 つを用意し、択一としている。キーワードには日本語の単語を用いているが、単語が複数の場合はそれらの単語に共通するイメージを連想できる画像群を選ぶよう指示している。また、回答者には伝えていないが A の画像群は日本語の単語から検索された画像で、B の画像群が英語の単語から検索された画像である。

多数決で “両方” が選ばれた場合は日英の単語から画像検索した結果に差がないため文化差無しと判断している。一方で、“A” もしくは “B” が選ばれた場合はどちらかの単語で検索された画像しか想起できなかったということなので文化差有りと判断している。もし、“該当無し” が最も多かった場合は、画像検索の結果が良くないことが考えられるため評価データから除外した。

3.3.2 最適値の検証

人手での判定結果を用いて、閾値の最適化をする。今回、人手での判定で文化差の有無を判定した概念 1000 個のうち、“文化差なし” の概念は 713 個、“文化差あり” の概念は 255 個、“該当なし” の概念は 32 個であった。このうち “該当なし” の概念 32 個を除いた 968 個の概念を用いて評価を行う。なお、このデータは文化差の有無に偏りのある不均衡データであるため、それぞれの 250 個ずつの合計 500 個をサンプリングする。サンプリングした 500 個のデータを 100 個ずつの 5 グループに分け、5 分割交差検証を行う。バリデーションデータとテストデータがそれぞれ 4:1 の比率になるように分割する。そして、分割されたデータがそれぞれ 1 回ずつテストデータになるように 5 回検証を繰り返し、それら 5 回のバリデーションとテストの結果を平均する。バリデーションデータでの結果から最適な閾値を見つけ、テストデータで見つけた閾値の評価を行う。バリデーションデータで提案手法の検出精度を求めた結果を図 7 に示す。

表 3: 5 分割交差検証による最適な閾値の検証

回数	閾値	Accuracy(%)		
		文化差なし	文化差あり	全体
1	0.55	78.0(39/50)	74.0(37/50)	76.0(76/100)
2	0.55	80.0(40/50)	78.0(39/50)	79.0(79/100)
3	0.55	72.0(36/50)	76.0(38/50)	74.0(74/100)
4	0.55	78.0(39/50)	72.0(36/50)	75.0(75/100)
5	0.55	76.0(38/50)	76.0(38/50)	76.0(76/100)
平均		76.8(38.4/50)	75.2(37.6/50)	76.0(76/100)

グラフの横軸は閾値、縦軸は正確さ (Accuracy) を表している。評価の結果、図 7 のように閾値が 0.55 の時、Accuracy は 76 % と最も高く、文化差検出のための最適な閾値は 0.55 である。表 3 は 5 分割交差検証の結果とその平均を示したものである。表 3 の閾値はバリデーションによって求められた最適な閾値

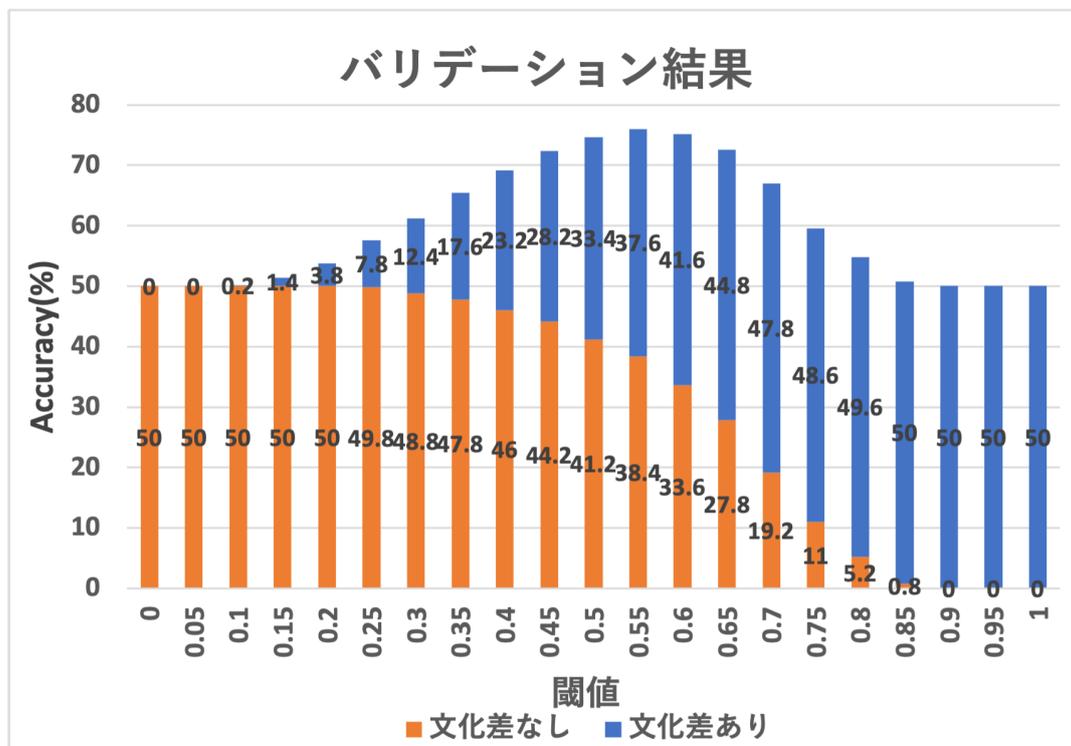


図 7: バリデーション結果

で, Accuracy は求めた閾値によってテストデータを文化差判定した結果である. それぞれ, 文化差なしに対する Accuracy, 文化差ありに対する Accuracy, 全体での Accuracy である. 平均した結果, 文化差なしの概念に対する Accuracy は 76.8 %, 文化差ありの概念に対する Accuracy は 75.2 %, 全体での Accuracy は 76 % である.

3.4 予備評価の分析

次に, 予備評価で使用した概念を分類し, 提案手法がどのような概念に効果的なのかを検証し, その結果をもとに評価データセットを作成する. WordNet では概念間の関係として上位概念 (Hypernym) と下位概念 (Hyponym) が提供されている. これらの関係性を用いて対象概念から上位概念を辿り, 各対象概念の Hypernym を調査する. Hypernym をもとに各概念を分類し, 種類ごとの文化差検出精度を調べる. 該当なしを除いた 968 個の概念の Hypernym を調査した結果, 949 個の概念の最上位概念が entity で, 残りの 15 個が固有名詞, 4 個が動詞であった. 本研究では大多数の entity に属する概念に限定して概念間の検出精度の差を分析する. 図 8 に示すように, entity の下位概念には physical entity と abstraction がある (表 4). 今回用いた 949 個の概念のうち 831 個は “physical entity” に, 118 個は “abstraction” に属している. “physical entity” は 831 個のうち “文化差あり” は 195 個であった. 一方, “abstraction” は 118 個のうち 47 個が “文化差あり” であった. “physical entity” と “abstraction” はそれぞれ “文化差あり” と “文化差なし” が同数になるようにアンダーサンプリングし, 閾値 0.55 で Accuracy を算出した. その結果, 表 5 に示すように, “abstraction” に比べて “physical entity” の方が僅かに Accuracy が高い. これは, 提案手法では画像から特徴ベクトルを抽出しているため, “physical entity” のような物理的に形を持ったもののほうが類似した特徴ベクトルを抽出でき, 文化差の検出精度が高まったと考えられる. 一方で, “abstraction” のように抽象的な概念では具体的な形が定まっていないため, 多様な特徴ベクトルが抽出され, 提案手法の検出精度が低くなったと考えられる.

次に, “physical entity” と “abstraction” の下位概念を分類する. 概念ごとの個数は “object” が 639 個, “matter” が 160 個, “process” が 5 個, “causal agent” が 16 個, “thing” が 11 個, “psychological feature” が 49 個, “attribute” が 23 個, “group” が 28 個, “communication” が 12 個, “measure” が 5 個, “relation” が 1 個

であった。各概念ごとに“文化差あり”と“文化差なし”の数が同数になるようにアンダーサンプリングし、Accuracyを算出した。表6の個数の部分がそれぞれサンプリング後のデータ数である。また、“relation”はデータ数が一つであったため、一つのデータの判定結果を参考値として載せている。Accuracyを確認した結果、“physical entity”の下位概念では、“process”や“thing”のAccuracyが低かった。一方で、“abstraction”の下位概念では、“attribute”、“group”のAccuracyが高いが、“psychological feature”や“communication”のAccuracyは低かった。上位概念ごとに分類したとき、各層ごとにそれぞれの概念によってAccuracyの差が多少見受けられた。そのため、概念の種類によって提案手法の性能の差があるかを調べるために、 χ^2 乗検定を行った(表7)。2層目の検定結果は $p=0.297965247$ 、3層目の検定結果は $p=0.445922655$ であり、どちらも有意な差(有意水準5%)が

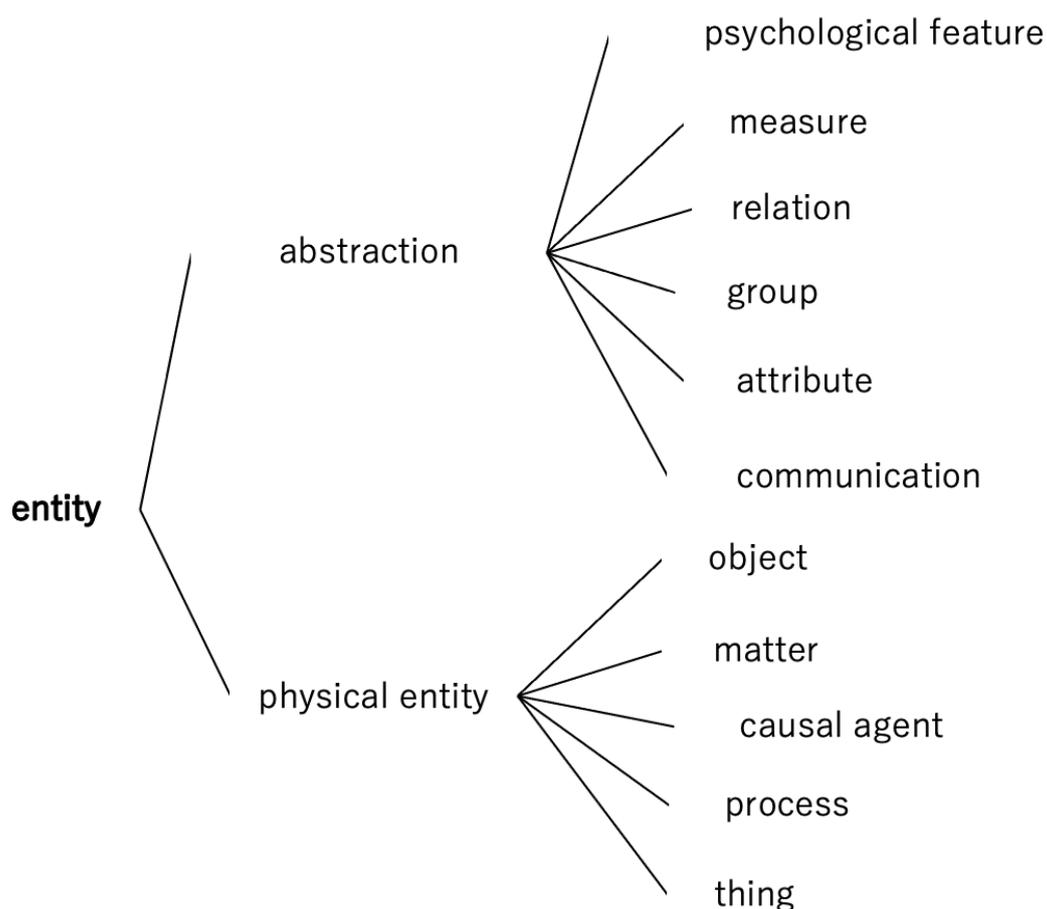


図8: 上位3層の Hypernym

表 4: 最上位の Hypernym

概念	説明
entity	(生命がある, あるいは生命がないに関わらずそれ自身の明確な存在を持つと感知される, 知られている, あるいは推定される何か that which is perceived or known or inferred to have its own distinct existence(living or nonliving)
physical entity	物理的な存在がある実体 an entity that has physical existence
abstraction	特定の例から共通点を抽出することによって形成された一般概念 a general concept formed by extracting common features from specific examples

表 5: 二層目の Hypernym ごとの Accuracy

概念	個数	正解	誤り	Accuracy(%)
physical entity	390	304	86	77.9
abstraction	94	71	23	75.5

認められなかった. このことから, 提案手法では, 今回のデータ数で検出できるような大きな性能差は概念の種類によって生じなかったと言える.

3.5 既知の文化差への適用

既存研究で報告されている文化差のある”リフォーム”, ”忍者”, ”noodle”, ”waterfall”という言葉 [4, 10, 17] に, 提案手法を適用し, 文化差の判定を行う. 表 8 に示すように “リフォーム”, “忍者” は提案手法によって文化差検出に成功したのに対して, “noodle”, “waterfall” は文化差の検出に失敗した. “リフォーム” の日本語では家の内装の画像が得られていたが, 英語では人の見た目の変化を表

表6: 三層目の Hypernym ごとの Accuracy

概念	個数	正解	誤り	Accuracy(%)
object	296	229	67	77.4
matter	68	48	20	70.6
process	2	1	1	50
causal agent	4	3	1	75
thing	8	5	3	62.5
psychological feature	48	33	15	68.8
attribute	12	10	2	83.3
group	18	16	2	88.9
communication	8	5	3	62.5
measure	4	3	1	75
relation	1	0	1	0

表7: 各層の検出精度の検定結果

層数	自由度	確率 (p)	χ^2 値
2層目	1	0.297965247	1.083278689
3層目	10	0.445922655	9.938304464

す画像が取得されていた。石田らの研究で報告されているように“リフォーム”は家の改築や模様替えの意味合いが強いが，“makeover”は化粧やダイエット後の変化を意味するため、文化差判定が成功したと考えられる。“忍者”は既存研究において具体的にどのような違いがあるのか説明がないため、取得された画像をもとに考察をする。日本語から取得された画像では忍者の絵や本などの画像が多いが、英語から取得された画像では忍者のコスプレをした人の画像が多かった。特に海外では、オリジナルの青い格好の忍者キャラの人気があり、そのコスプレなどが流行っているため、取得された画像に差があり、判定に成功したのではないかと考えられる。“noodle”や“waterfall”は柳井らの研究で発見されている文化差である。“noodle”はヨーロッパではスパゲッティが最も想起されるのに対して、日本ではラーメンが最も多く想起されると明らかにしている。どちら

表 8: 既知の文化差と提案手法での判定結果

日本語	英語	類似度
“リフォーム”	“makeover”	0.195
“忍者, 忍の者, 忍びの者”	“ninja”	0.351
“ヌードル, 麵”	“noodle”	0.585
“瀧, 落水, 垂水, 飛瀑, 飛泉, 滝, 瀑布, 水簾”	“waterfall, falls”	0.779

も同じ麺類であるため、提案手法では類似度が高くなり、文化差なし判定になったと考えられる。同様に “waterfall” はアジアでは美しい滝を想起するが、南アメリカでは力強い滝を想起するため、文化差があると述べられている。このように感じ方は異なるがどちらも同じ滝が想起されているため、提案手法では文化差なし判定になったと考えられる。

第4章 評価データセット

文化差検出手法を評価するための評価データセットの作成手法について説明する。今回、評価データセットでは日本語と英語、インドネシア語の3言語でアンケートを作成し、その結果をもとに、日英間の文化差、日本とインドネシア語間の文化差を判定した2つのデータセットを作成した。これらは両方同じ方法でデータセットを作成しているため、日英での評価データセットをもとに評価データセットの作成方法を説明する。

4.1 評価データセットの作成

まず、予備評価で用いたデータの問題点を挙げる。

- 英語話者の視点が入っていない
- 画像単体の評価を得ることが難しいため、文化差の判定が
- 同音異義語を考慮できていない
- 概念の種類に偏りが見られる

予備評価で使用したデータでは、日本語から取得した画像と英語の単語で取得した画像をもとに日本語話者のみが文化差判定を行っている。そのため、英語話者の視点や言葉といった要素が含まれていない。また、予備評価のアンケートでは日本語の単語と画像群2つ(英語と日本語)を見せ、どちらの画像が単語からより想起できるかを質問している。この質問では、画像に共通しているものから概念が予想でき、それをもとに単語から想起できると答えることが可能であるため、その画像単体で単語から想起できるかがわかりづらい。また、単語と画像しか与えていないため、同音異義語を考慮できていない。例えば、“トラック”のように車両としてのトラックと陸上の競技場で使用されるトラックがある場合、単語だけではどちらも想起できてしまう。従って、画像だけではなく単語や説明といった複数の要素をもとに日英の両方の視点から対象概念の文化差判定を行なった評価データセットを作成する必要がある。このような評価データセットを作成するために概念を表す画像と言葉といった要素を日英の両言語話者の視点から文化差を判定することで概念間の文化差の有無を表した評価データセットを作成する。

具体的には、概念辞書から概念を1000個取り出し、各概念ごとに日英の単語群、概念の説明、単語群から画像検索によって取得された画像を日英の言語話者

に見せる。各話者は母言語の情報をもとに取得されたそれぞれの画像が単語や説明からイメージできるかを5段階で評価する。言語によってこれらの概念の評価に差が見られる概念を文化差あり、差が見られない概念を文化差なしとしてラベリングする。

まず、WordNetの上位三層目の概念(図8)をもとにSynset(概念)を1000個選り出す。予備評価ではデータをWordNetのHypernymに基づいて分類し、それぞれのAccuracyを評価した。それらの結果を統計検定を用いてHypernymの種類ごとにAccuracyの差があるかを評価した結果、有意水準5%未満で有意差はなかった。しかし、予備評価で用いたデータは分類したHypernymの個数がそれぞれ大きく違う。加えて、有意差はなかったもののAccuracyの差が見受けられるため、評価データセットではこれらの分類したHypernymごとの個数がなるべく等しくなるようにする。そのため、選り出した1000個のSynsetは上位三層目の概念(psychological feature~thing)の下位概念がなるべく均等になるように選り出す。

次に、WordNetと日本語WordNet(インドネシア語の場合、インドネシア語WordNet)から英語と日本語(インドネシア語)の単語群を取り出す。表1のようにそれぞれのSynsetには各言語でそのSynsetを表す単語群と説明が付けられている。このうちの単語群を用いて各言語ごとに画像検索を行う。単語が複数の場合、AND検索を用いることでそのSynsetを表す画像を取得する。取得された画像のうち、上位10件ずつの画像を用いる。

各Synsetごとの日英インドネシア語の単語群とそれらから取得された画像群、加えてSynsetの説明を用いてアンケートを作成する。そして、各言語話者に日英インドネシアの各画像群がどれほど説明やキーワードから想起できるかを問う。そして、各言語ごとに回答を集計・数値の平均を行い、日本語話者、英語話者、インドネシア語話者の結果をまとめる。その結果をもとに各言語話者の回答がそれぞれ日本語の画像群、英語の画像群、インドネシア語の画像群が想起できるか、想起できないかを一意に決める。

このうち、日本語話者と英語話者の日英の画像が想起できるかの結果をもとに、言語間で判定に違いが現れたSynsetを文化差ありと判定する。一方で、言語間で判定に違いがないSynsetを文化差なしとする。また、画像検索の結果が悪い場合、そのSynsetは対象外としてデータセットから取り除く。

4.2 アンケート

アンケートの質問は大きく以下の5種類である。

- A 画像に写っているもの同士が同じものを指しているか (1~5 段階)
- B 説明と画像がどのくらいあっているか (1~5 段階)
- C キーワードと画像がどのくらい一致しているか (1~5 段階)
- D 説明とキーワードがどのくらい一致しているか (1~5 段階)
- E キーワードと説明をみて、この概念を知っているかどうか (Yes or No)

また、5段階評価の尺度を以下の通りに示している。

1. 全く違う
2. 少しはあっている
3. ある程度あっている
4. おおかたあっている
5. 完璧にあっている

質問のうち、D(説明とキーワードがどのくらい一致しているか)とE(キーワードと説明をみて、この概念を知っているかどうか)においては、Synsetにつき質問が一回である。一方で、A(画像に写っているもの同士が同じものを指しているか)、B(説明と画像がどのくらいあっているか)、C(キーワードと画像がどのくらい一致しているか)の質問はそれぞれの画像に合わせて答える必要がある。このように、各画像群別に質問を分けることによって、各画像群の単体でのキーワードや説明から想起できるかという評価を得ることができる。それらの表組みが表9である。実際のアンケートは図9である。

まず、各言語で取得した画像群と説明文、キーワードを回答者に情報として与える。回答者にどの言語で取得した画像かを不明確にするために画像群を並べる順番はSynsetごとに変更している。説明文とキーワードはアンケート回答者の母語言語に合わせて、それぞれのWordNet上から取り出したSynsetの説明と単語群を用いている。これらをもとに表9内の*のついている部分を回答者に答えてもらう。

それぞれの質問では表9の(1)~(3)の部分で質問Aについて問う。(1)では画像群1と画像群2がどの程度同じものを指しているか、(2)では画像群2と画像群3、(3)では画像群1と画像群3について質問Aを問う。(4)~(6)では説明文と各画像群をもとに質問Bを問う。(7)~(9)の部分ではキーワードと各画像群に

表 9: アンケートの表組み

画像	質問 A	質問 B	質問 C	質問 D	質問 E
		説明文	キーワード		
画像群 1	*(1)	*(4)	*(7)		
画像群 2	*(2)	*(5)	*(8)	*(10)	*(11)
画像群 3	*(3)	*(6)	*(9)		

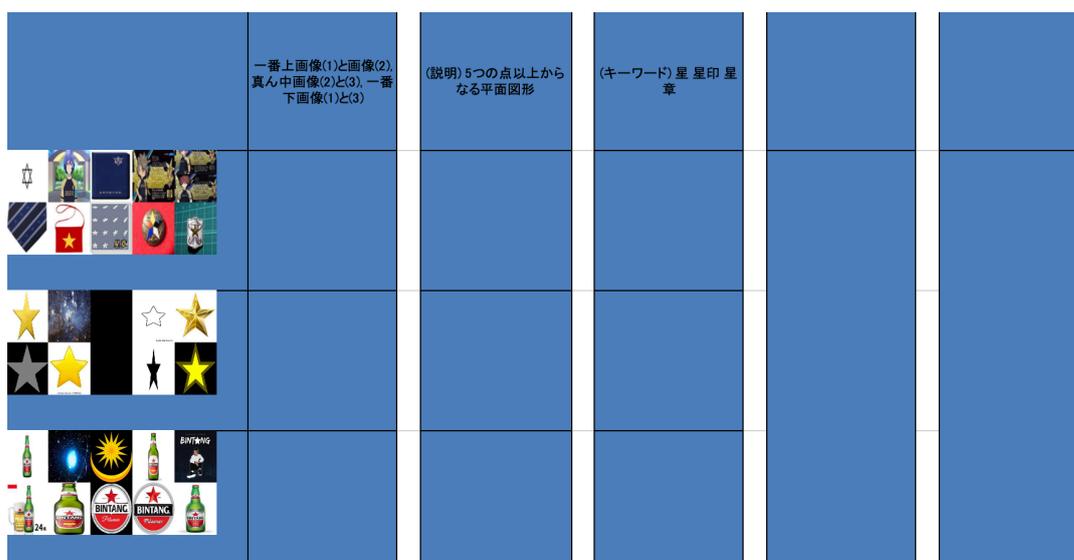


図 9: 日本人アンケート

もとに質問 C を問う. (10) と (11) では説明文とキーワードをもとにそれぞれ質問 D と質問 E を問う.

4.3 評価データセットの集計

次に, これらのアンケートをもとに評価データセットを集計する. まず, 各言語話者ごとにアンケート結果を取りまとめる. 今回, 日本語話者 5 人と, 英語話者 1 人, インドネシア語話者 2 人にアンケートを回答してもらった. このうち日本語話者 5 人で表 9 の (1) の値を平均する. このとき, (11) の回答が “No” になっている回答者のアンケート結果は含めない. (1)~(10) の平均を算出し, 日本人のこの Synset に対するアンケート結果を一意に決定する. このように各言語話者ごとにアンケート結果を一意に決定する.

日英間では日本語で取得された画像と英語で取得された画像群を対象にアン

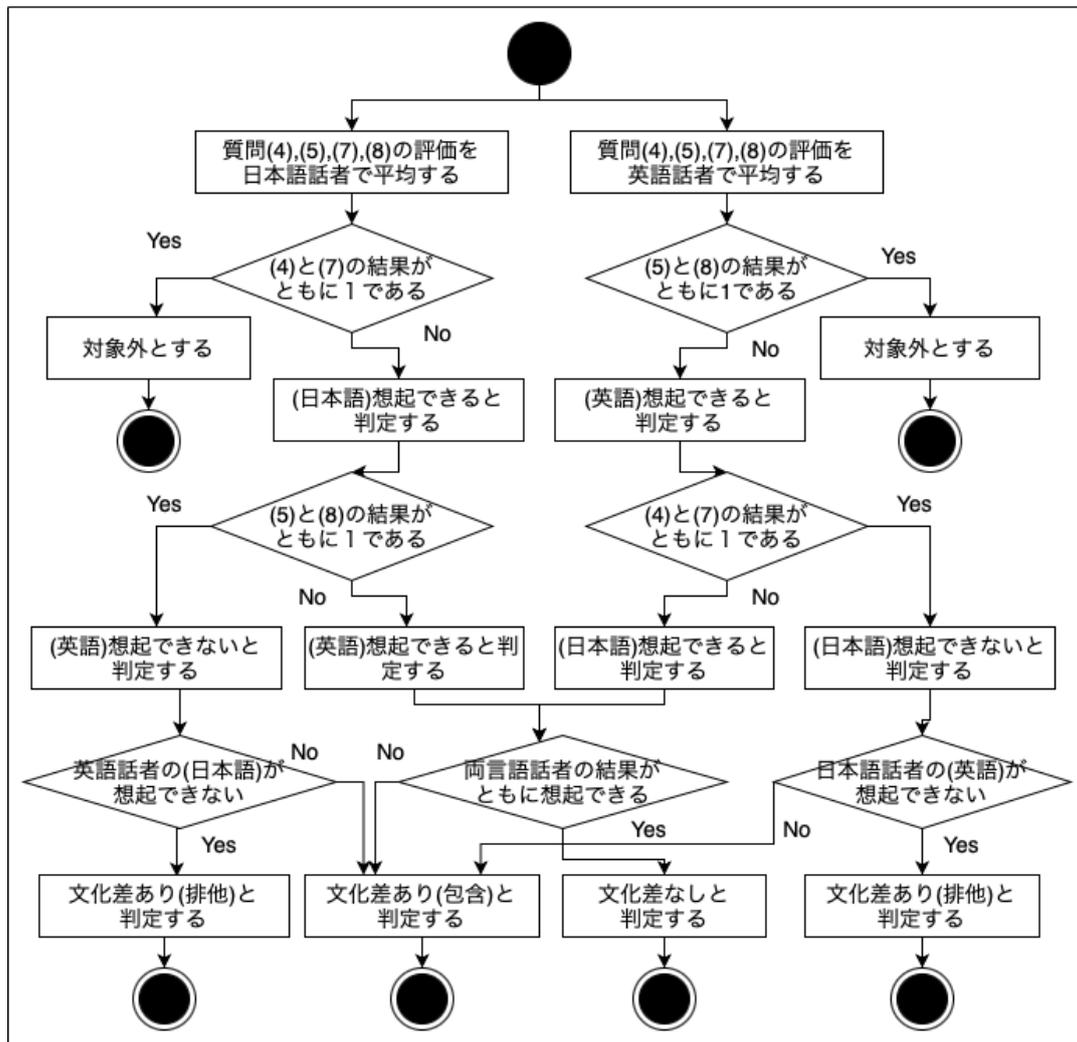


図 10: 集計の手順

表 10: 文化差なし

	日本語話者	英語話者
日本語の画像	想起できる	想起できる
英語の画像	想起できる	想起できる

ケートから情報を抽出し、日本語話者と英語話者のアンケート結果を比較して文化差の有無を判定する。図9のSynsetでは一番上の画像群1が日本語のキーワードで取得された画像群で、真ん中の画像群2が英語のキーワードで取得された画像群、そして、一番下の画像群3がインドネシア語のキーワードから取得

表 11: 文化差あり (包含関係)

	日本語話者	英語話者
日本語の画像	想起できる	想起できる
英語の画像	想起できない	想起できる

表 12: 文化差あり (排他関係)

	日本語話者	英語話者
日本語の画像	想起できる	想起できない
英語の画像	想起できない	想起できる

された画像群である。そのため、日英間では図 9 の場合、表 9 の (4), (5), (7), (8) の値を参照する。

まず、日本人のアンケート結果の場合、(4) と (7) の結果が両方とも “1” であるかどうかを確認する。画像群 1 は日本語のキーワードから取得された画像群であるため、日本人にとってこれらの画像群がキーワードと説明の両方から想起できないということは画像検索の結果が悪いと言える。そのため、(4) と (7) の結果が “1” の場合、“対象外” としてデータセットからこの Synset を外す。もし、どちらか一方でも “1” 以外の場合、日本語の画像は “想起できる” と判定する。次に、(5) と (8) は英語のキーワードから取得された画像群である。これらの結果がともに “1” である場合、日本語話者にとって英語の画像は “想起できない” と判定する。どちらか一方でも “1” 以外である場合、英語の画像は “想起できる” と判定する。このように、日本語話者にとって、日本語の画像群と英語の画像群のアンケート結果が “想起できる” か “想起できない” かをそれぞれ決定する。

同様に英語話者の場合、(5) と (8) の値が両方とも “1” のとき “対象外” としてデータセットから Synset を外す。次に、(4) と (7) の結果がともに “1” の場合に “想起できない”、どちらか一方でも “1” ではない場合に “想起できる” と判定し、英語話者の日本語の画像群と英語の画像群の結果を決定する。

そして、日英の両言語話者の 2 つの結果を比較し、その結果が一致している場合にその Synset は “文化差なし” と判定する (表 10)。一方で、表 11 のように日本語話者は日本語の画像も英語の画像もともに “想起できる” に対して、英語話者は日本語の画像が “想起できない” である場合、この Synset は文化差が存在す

表 13: 日英間の評価データセットの内訳

概念	synset の個数	文化差あり		文化差なし	対象外
		排除	包含		
object	101	6	10	74	11
matter	102	4	5	93	0
process	90	11	16	46	17
causal agent	97	5	17	66	9
thing	96	2	23	59	12
psychological feature	98	9	22	47	20
attribute	35	5	5	14	11
group	95	6	17	52	20
communication	96	11	29	44	12
measure	96	6	17	37	36
relation	94	6	25	33	58
合計	1000	71	186	565	178

る. この場合, 日本語話者にとっては両方想起できるが, 英語話者にとっては片方しか想起できないため, 日本語話者が考える想起イメージの方が概念の範囲が広い. そのため, 2.2 章で説明した“包含関係”に該当するため, “文化差あり (包含関係)”と判定する. そして, 表 12 のように日本語話者も英語話者をお互いの相手の言語で取得した画像群を“想起できない”場合, どちらにも共通しない概念の範囲が存在するため, 2.2 章で説明した“排除関係”に該当する. よって, “文化差あり (排除関係)”と判定する.

このようにアンケートから各 Synset を“文化差なし”, “文化差あり (包含関係)”, “文化差あり (排除関係)”, “対象外”と判定する. このうち, “対象外”の Synset を取り除き, 残った Synset にそれぞれの判定結果をラベル付けする. これを本手法を評価するための評価データセットとする. 作成された評価データセットの内訳が図 13 である.

まず, 1000 個の概念のうちから“対象外”の 178 個の概念を取り除いた計 822 個のデータが出来上がった. 内訳としては“文化差あり”の概念が 257 個で, そのうち 71 個が排除関係の文化差, 残り 186 個包含関係の文化差であった. そし

表 14: 日インドネシア間の評価データセットの内訳

概念	synset の個数	文化差あり		文化差なし	対象外
		排他	包含		
object	101	4	10	78	9
matter	102	2	2	93	5
process	90	9	5	69	7
causal agent	97	5	11	70	11
thing	96	5	12	73	6
psychological feature	98	5	17	64	12
attribute	35	3	6	20	6
group	95	8	9	61	17
communication	96	3	11	71	11
measure	96	8	12	41	35
relation	94	7	23	43	21
合計	1000	59	118	683	140

て, “文化差なし” の概念が 565 個であった. また, 3 層目の概念の下位概念をなるべく均等になるように Synset を選択したが, “attribute” の下位概念は 35 個しか日本語が対応づけられていなかった. そのため, “attribute” のみ数が少なくなっている.

同様の手法で日本とインドネシア語間の評価データセットを作成した. 日インドネシア間では, 日本語とインドネシア語話者のアンケート集計結果から日本語の画像とインドネシア語の画像に対応した質問から結果を取り文化差の有無のラベル付けを行う. それによって, 作成された日インドネシア間での評価データセットの内訳が表 14 である. まず, 1000 個の概念のうちから “対象外” の 140 個の概念を取り除いた計 860 個のデータが出来上がった. 内訳としては “文化差あり” の概念が 177 個で, そのうち 59 個が排他関係の文化差, 残りの 118 個の包含関係の文化差であった. そして, “文化差なし” の概念が 683 個であった.

第5章 異常値検出を用いた改善

異常値検出を用いた改善方法を説明する。本手法では各言語ごとに10枚の画像の特徴ベクトルを単純平均によって、対象概念を表す平均特徴ベクトルを算出し、それら平均特徴ベクトル間の類似度に基づいて文化差検出している。現在の問題として、10枚の画像の中に対象概念と異なる全く関係のない画像が含まれていた場合、正確に文化差検出できない可能性がある。そのため、これらのノイズとなる画像を異常値検出を用いて取り除き、純粋な対象概念を表す平均特徴ベクトルを作ることによって Accuracy を向上する。

具体的には、各単語群から画像検索を行い、上位10件の画像の特徴ベクトルを取得する。ここから従来の提案手法と異なった改善手法を提案する(図11)。まず、それぞれ10枚の特徴ベクトルに異常値検出を行う。その結果に基づき特徴ベクトルに重み付けを行い、加重平均により平均特徴ベクトルを合成する。その後、従来手法と同様に類似度を算出し、閾値を用いて文化差判定を行う。異常値検出の手法は3つの手法を試したため、これら3つの手法を説明する。

5.1 K-means

異常値検出の手法の1つ目はK-meansを用いた手法である。K-meansとは非階層クラスタリングのアルゴリズムで、これを用いて10個の特徴ベクトルをクラスタリングする。VGG16では本来特徴ベクトルを算出し、それらのベクトルから画像認識を行うことができる。そのため、画像が異なっても同じものが写った画像では近似した特徴ベクトルが生成されると考えられる。そのため、10枚の画像特徴ベクトルをK-meansでクラスタリングしたときにターゲットと異なる画像が含まれていた場合、ノイズとなる画像はより少数のクラスタに分類されると考えられる。これを用いて、多数派のクラスタに分類されたベクトルの重みを高く付ける。一方で、少数派のクラスタに分類されたベクトルの重みを小さくする。

具体的に、10個の画像特徴ベクトルにK-meansを用いてクラスタリングを行い、各クラスタに分類された要素数に応じて、特徴ベクトルに重み付けを行う。重みの値はそのベクトルが分類されたクラスタの要素の数と同じ重みを付与する。例えば、10個の特徴ベクトルのうち、7個のベクトルがクラスタAに、残り3個のベクトルがクラスタBに分類された場合、クラスタAの特徴ベクトル全て

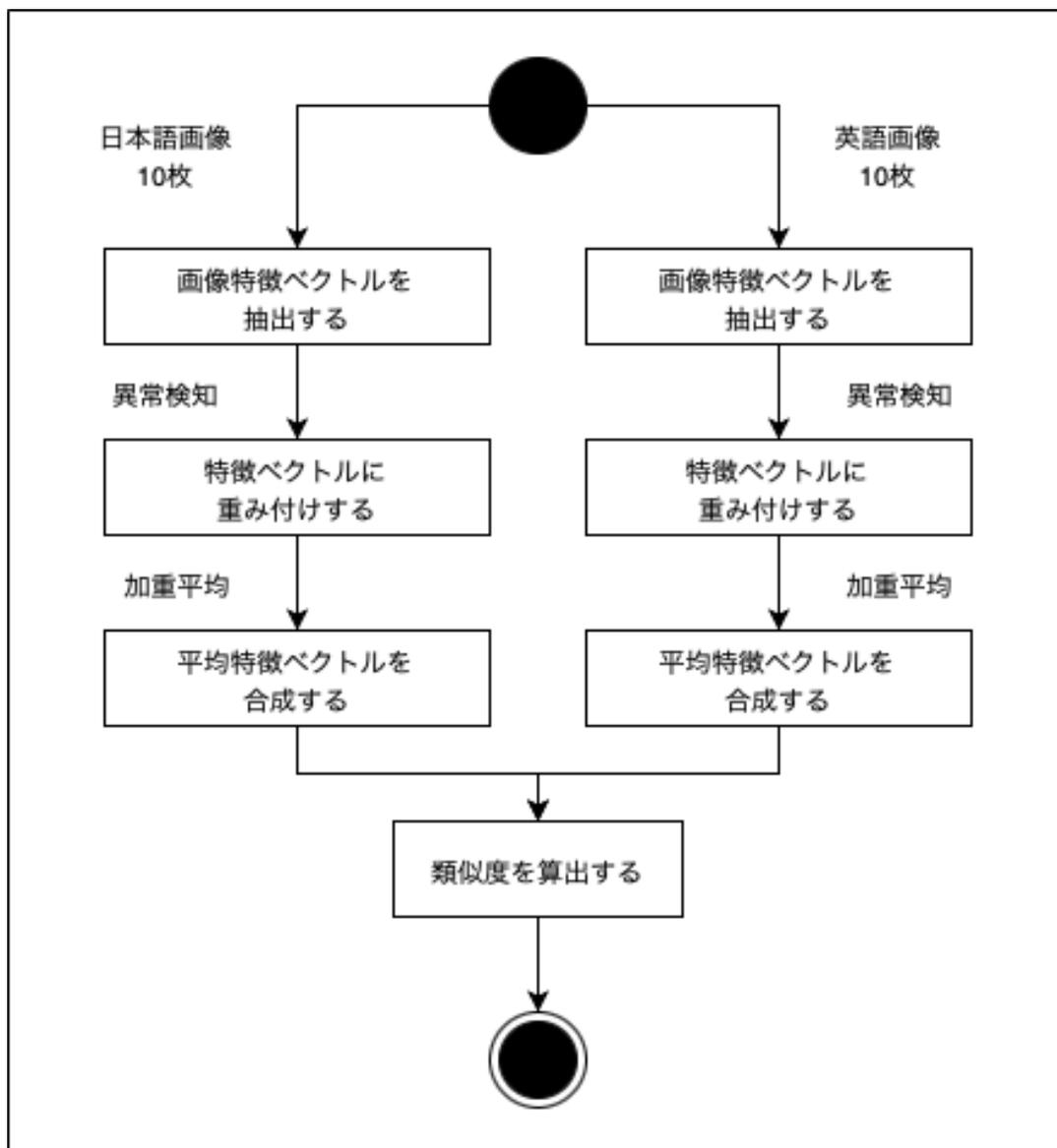


図 11: 改善手法

に重み“7”を付与し、クラスタ B の全ての特徴ベクトルに重み“3”を付与する。

また、クラスタリングする特徴ベクトルが 10 個と少ないため、K-means で分類するクラスタの数は 2 つとする。

5.2 DBSCAN

2 つ目の異常値検出手法は DBSCAN を用いた手法である。DBSCAN とはデータの密度を用いてクラスタリングを行う手法で、クラスタと見なすための最小

の数と距離を初期値として与えることでクラスタリングを行う。K-means ではデータを決められたクラスタの数にクラスタリングするため、異常値を検出することができない。そのため、複数のノイズ画像が含まれていた場合、全てのノイズデータを少数派のクラスタに分類できない可能性がある。一方で、DBSCAN では与えられた初期値をもとに密度でクラスタリングを行うため、クラスタリング可能なものをクラスタリングし、クラスタリングできないデータは異常値として検出することができる。そのため、DBSCAN を用いて各特徴ベクトルをクラスタリングし、異常値として検出されたデータの重みを限りなく低くし、残りのクラスタリングできたデータをクラスタ内の要素数に基づいて重み付けを行う。

具体的には、10 個の特徴ベクトルに DBSCAN を用いてクラスタリングを行う。この結果、異常値とみなされた特徴ベクトルの重みは“1”に設定し、残りの特徴ベクトルの重みは K-means 同様に分類されたクラスタ内の要素数と同じ数を付与する。

また、各言語ごとに特徴ベクトルの数が 10 個と少ないため、クラスタと見なすための最小の数は 2 個としている。また、その距離は事前に 10 個の特徴ベクトルを DBSCAN でクラスタリングした際にクラスタが生成された距離の値:1500 としている。また、距離の値を 1000 以下にするとクラスタは生成されず、全てのベクトルが孤立した状態となる。一方、距離の値を 2000 以上にした場合、ノイズデータを含めたデータでも全てのベクトルが同一のクラスタに分類されたため、クラスタが生成され、異常値も検出できる 1500 としている。

5.3 IsolationForest

3 つ目の異常値検出手法は IsolationForest を用いた手法である。IsolationForest は RandomForest のように決定木を用いた異常検知手法であり、DBSCAN や K-means とは異なりクラスタリングするのではなく異常値を検出するだけである。これを用いて画像の特徴ベクトルから異常値を検出し、重み付けを行う。

具体的には、10 個の特徴ベクトルに IsolationForest を用いて異常検知を行う。もし、ノイズ画像のような異常値が含まれている場合、これによって異常値を検出し、その異常値の重みを“1”にする。残りの特徴ベクトルにベクトルの総数“10”から異常値の数を引いた数を重みとして付与し、これらを用いて加重平均を行う。

第6章 評価実験

本研究を評価するための評価実験について説明する。

6.1 実験の手順

評価実験の手順について説明する。評価実験には作成した評価データセットを活用する。評価データセットを使用して、予備評価と同様に Accuracy を評価指標として提案手法を評価する。

まず、作成した評価データセットは文化差ありデータが 257 個、文化差なしデータが 565 個の不均衡データである。そのため、これらを文化差ありデータと文化差なしデータが同数になるようサンプリングし、正確な Accuracy を計測する。これによって、文化差ありデータと文化差なしデータが 250:250 の計 500 個になるようにサンプリングし、評価実験に使用する。次に、これらの 500 個のデータを 100 個ずつの 5 つのグループに分ける。この 5 つのデータグループをバリデーションデータとテストデータが 4:1 になるように割り振る。データの偏りを防ぐため、5 分割交差検証によって全てのデータグループがテストデータになるようにデータグループの組み合わせを変えながら 5 回バリデーションとテストを繰り返す。最後に、バリデーションとテストのそれぞれ 5 回分の結果を平均し、最終的なテスト結果を算出する。また、算出される類似度から文化差を判定するための閾値は予備評価と同様に 0.0~1.0 まで 0.05 おきに変動させ、最も Accuracy の高い閾値を評価に採用する。

また、従来の単純平均によって平均特徴ベクトルを合成する方法で評価を行なった後、各手法を用いて加重平均により平均特徴ベクトルを合成する方法で Accuracy を検証し、従来手法と比較する。これにより従来手法と比べてどの程度 Accuracy が向上するかを評価する。また、比較するために使用するデータは同様の 500 個のデータを使用する。

6.2 実験結果

次に、実験の結果を示す。最初に、日英間での文化差判定の結果である。図 12 が従来の単純平均によって、平均特徴ベクトルを生成する手法でテストを行なった結果である。この図 12 は 6.1 章で説明したように文化差ありデータと文化差なしデータが 250:250 の合計 500 個のデータを用いて、5 分割交差検証を行なっ

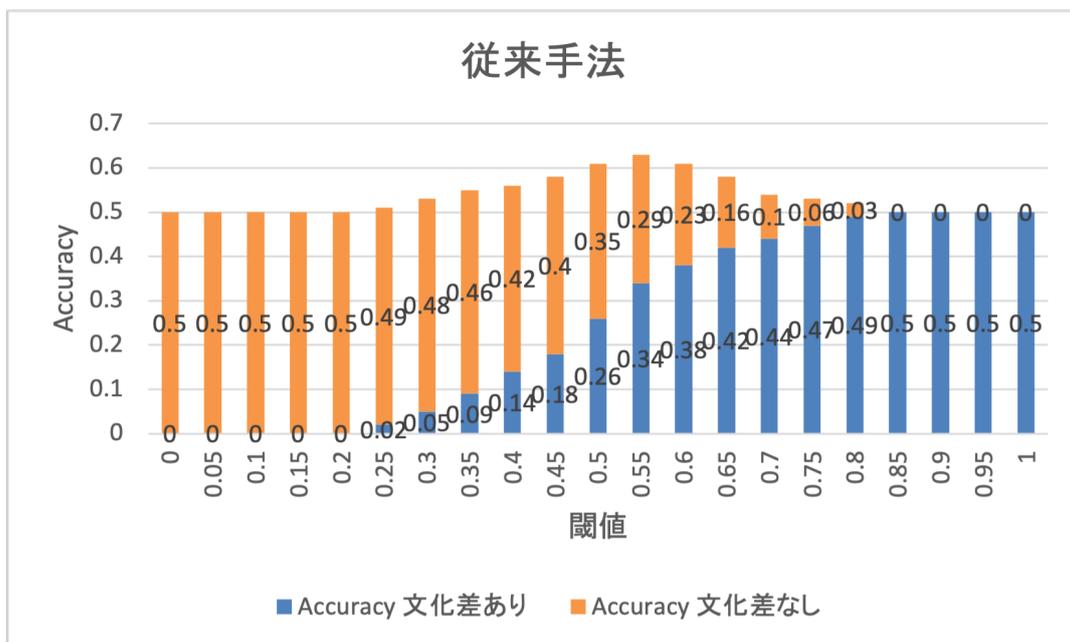


図 12: 従来手法でのテスト結果

た際のテストデータの結果である。横軸は閾値を示し、縦軸は Accuracy を示している (3.2 章, 数式 1 参照)。また、青色の部分には Accuracy のうち、文化差ありデータに対する Accuracy を示し、オレンジ色の部分には文化差なしデータに対する Accuracy を示している。このグラフは閾値を 0.0 から 1.0 まで変動させた際に、本手法がどの程度 Accuracy があるかを評価したグラフである。なお、閾値 0.0 の時は全ての類似度を“文化差なし”と判定しているため、文化差なしデータを全て正確に文化差判定できている。一方で、文化差ありデータは全て文化差判定に失敗している。逆に閾値 1.0 に設定した場合、全てのデータに対して“文化差あり”と判定しているため、文化差ありデータ 50 個は全て正確に文化差判定できているのに対し、文化差なしデータは全て文化差判定に失敗している。そのため、両端の閾値 0.0 と閾値 1.0 の時はともに全体の Accuracy は 0.5(50%) である。

では、図 12 の結果、最も Accuracy が高くなるのは閾値 0.55 に設定した時である。その際の Accuracy は 0.63(63%) である。また、文化差ありデータに対する Accuracy は 0.34 の 2 倍で 0.68(68%) である。一方で、文化差なしデータに対する Accuracy は 0.29 の 2 倍で 0.58(58%) であった。なお、これらの文化差ありデータと文化差なしデータに対する Accuracy は図 12 内ではデータ全体のうち、

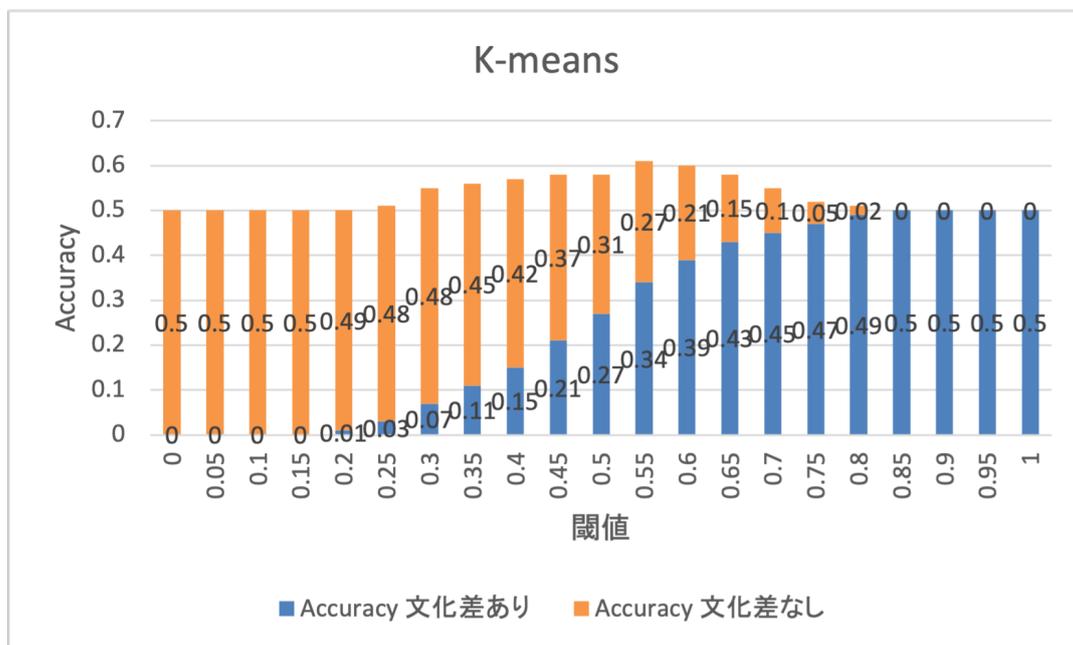


図 13: K-means を用いた手法でのテスト結果

半分ずつしかそれらのデータがない。文化差ありデータのみに対する Accuracy を表す場合、2 倍する必要がある。

次に、5.1 章で説明した K-means を用いてクラスタリングを行い、分類されたクラスタの要素数に応じて重みを用いて平均特徴ベクトルを生成する手法の結果である。また、先程の 12 と比較するためにデータは同じ 500 個のデータを使用している。図 13 のように K-means を用いた手法では Accuracy が最も高くなったとき、従来手法と同様に閾値は 0.55 であった。その時の Accuracy が 0.61(61 %) であった。また、文化差ありデータに対する Accuracy は 0.68(68 %) で、文化差なしデータに対する Accuracy は 0.54(54 %) であった。

次に、5.2 章で説明した dbscan を用いてクラスタリングを行い、クラスタ内の要素数に応じた重みを持った特徴ベクトルで平均特徴ベクトルを生成する手法の結果である。DBSCAN を用いた手法では、閾値 0.5 と閾値 0.55 の時に Accuracy が最も高くなっている。この時の Accuracy が 0.59(59 %) であり、文化差ありデータに対する Accuracy は閾値 0.5 の時に 0.54(54 %)、閾値 0.55 の時に 0.68(68 %) であった。一方で文化差なしデータに対する Accuracy は閾値 0.5 の時に 0.64(64 %) で、閾値 0.55 の時に 0.50(50 %) であった。

最後に、5.3 章で説明した IsolationForest を用いて異常値検出を行い、異常値が

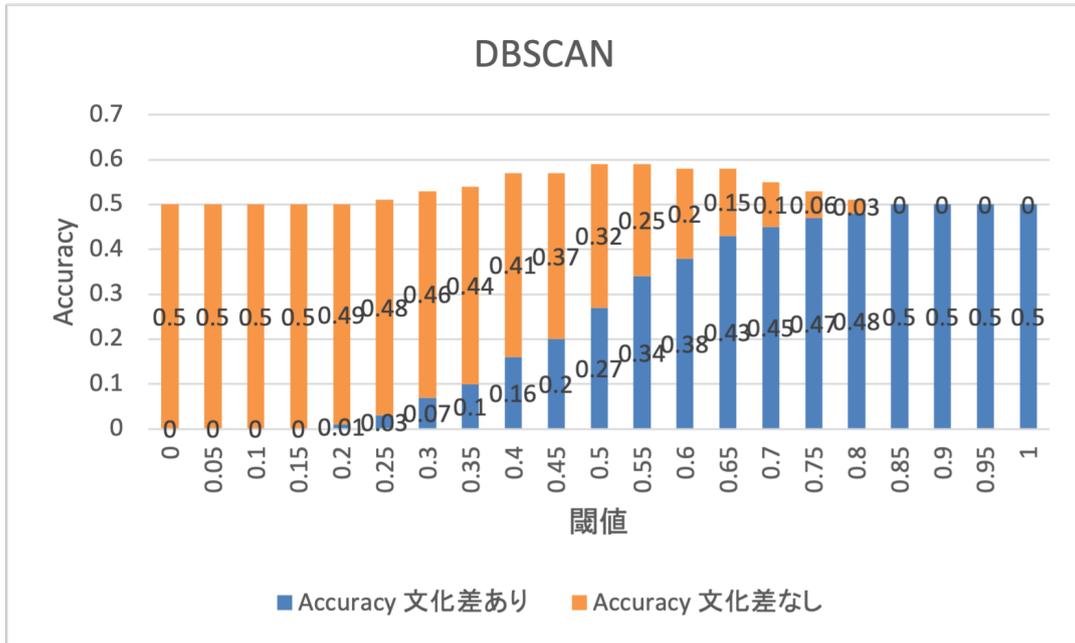


図 14: DBSCAN を用いた手法でのテスト結果

表 15: (日英間) 各手法での比較

手法	閾値	Accuracy(%)		
		文化差あり	文化差なし	全体
従来手法	0.55	68.0(34/50)	58.0(29/50)	63.0(63/100)
K-means	0.55	68.0(34/50)	54.0(27/50)	61.0(61/100)
DBSCAN	0.5	54.0(27/50)	64.0(32/50)	59.0(59/100)
	0.55	68.0(34/50)	50.0(25/50)	
IsolationForest	0.55	68.0(34/50)	56.0(28/50)	62.0(62/100)

あった場合にその特徴ベクトルの重みを低くする手法の結果である (図 15)。この IsolationForest を用いた手法では閾値 0.55 の時に Accuracy が 0.62(62%) で最も高かった。そのうち文化差ありデータに対する Accuracy は 0.68(68%) で、文化差なしデータに対する Accuracy は 0.56(56%) であった。

表 15 はそれぞれの手法の Accuracy を比較したものである。最も Accuracy が最も高かったのが従来手法であり、その他の異常値検出を用いた改善手法は従来手法と比較して Accuracy は向上しなかった。

次に、日インドネシア間での文化差判定の結果である。

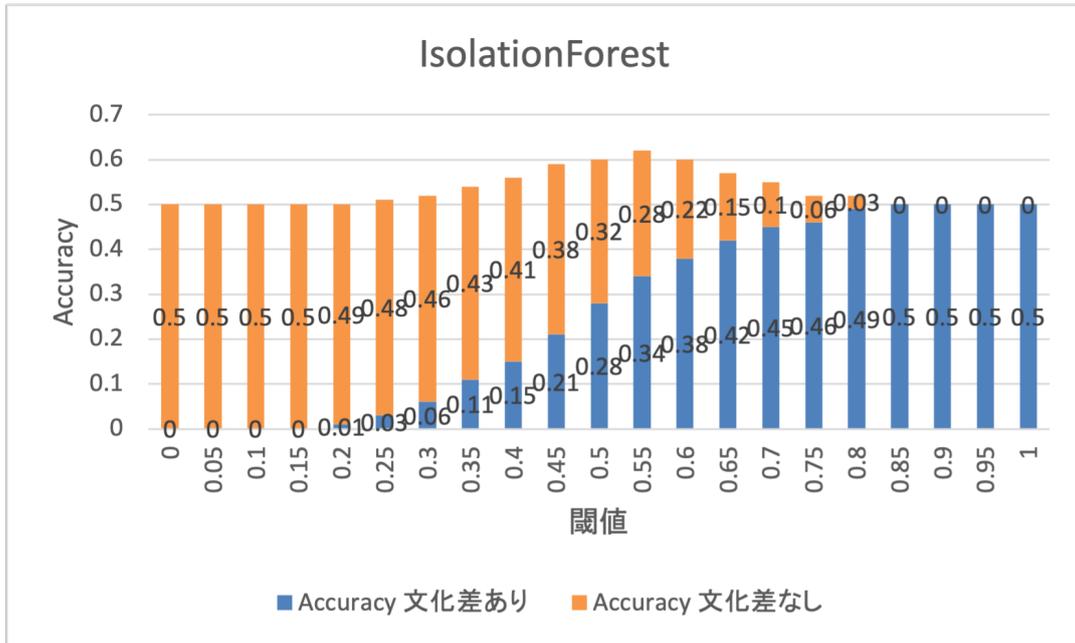


図 15: IsolationForest を用いた手法でのテスト結果

表 16: (日インドネシア間) 各手法での比較

手法	閾値	Accuracy(%)		
		文化差あり	文化差なし	全体
従来手法	0.4	40.0(20/50)	84.0(42/50)	62.0(62/100)
	0.45	54.0(27/50)	70.0(35/50)	
K-means	0.45	56.0(28/50)	68.0(34/50)	62.0(62/100)
DBSCAN	0.35	34.0(17/50)	84.0(42/50)	59.0(59/100)
	0.45	54.0(27/50)	64.0(32/50)	
	0.5	66.0(33/50)	52.0(26/50)	
IsolationForest	0.4	44.0(22/50)	80.0(40/50)	62.0(62/100)

まず、従来手法によるテスト結果である。日本語とインドネシア語間では、1000件のうち、文化差ありデータが177個、文化差なしデータが683個であったため、日英間での文化差判定と同様に文化差ありと文化差なしが同数になるようにサンプリングを行う。これにより、文化差ありと文化差なしが150:150の均衡データにしたのち、5分割交差検証によりデータを60個ずつの5グループに分け評価を行う。図16が日インドネシア間での従来手法の文化差判定結果である。こ

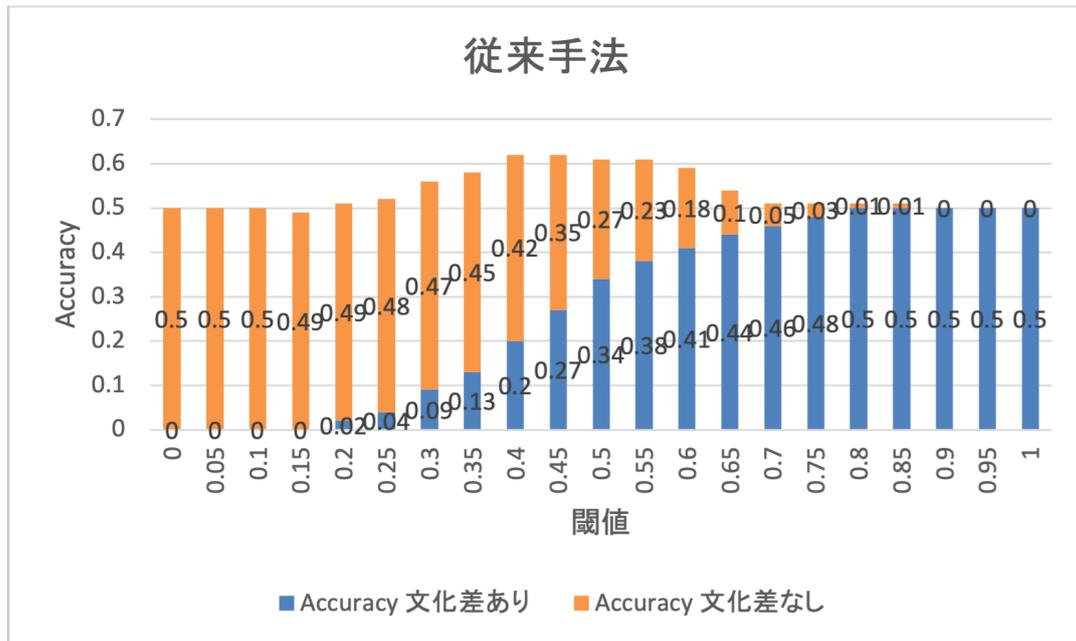


図 16: 従来手法結果 (日インドネシア)

の結果, Accuracy は 0.62(62%) で閾値 0.4 と閾値 0.45 の時に最大となった。また, 文化差ありデータの Accuracy は閾値 0.4 の時に 0.4(40%), 閾値 0.45 の時に 0.54(54%) であった。一方で, 文化差なしデータの Accuracy は閾値 0.4 の時に 0.84(84%), 閾値 0.45 の時に 0.7(70%) であった。

続いて, 図 17 は K-means を用いた手法での日インドネシア間での結果である。この結果から閾値 0.45 の時に Accuracy が 0.62(62%) と最も高かった。この時の文化差ありデータの Accuracy が 0.56(56%) で, 文化差なしデータの Accuracy が 0.68(68%) であった。

そして, 図 18 は DBSCAN を用いた手法での日インドネシア間での結果である。この結果, Accuracy は 0.59(59%) で閾値 0.35 と閾値 0.45, 0.5 の時に最大となった。また, 文化差ありデータの Accuracy は閾値 0.35 の時に 0.34(34%), 閾値 0.45 の時に 0.54(54%), 閾値 0.5 の時に 0.66(66%) であった。一方で, 文化差なしデータの Accuracy は閾値 0.35 の時に 0.84(84%), 閾値 0.45 の時に 0.64(64%), 閾値 0.5 の時に 0.52(52%) であった。

最後に, 図 19 は IsolationForest を用いた手法での日インドネシア間での結果である。この結果から閾値 0.4 の時に Accuracy が 0.62(62%) と最も高かった。この時の文化差ありデータの Accuracy が 0.44(44%) で, 文化差なしデータの

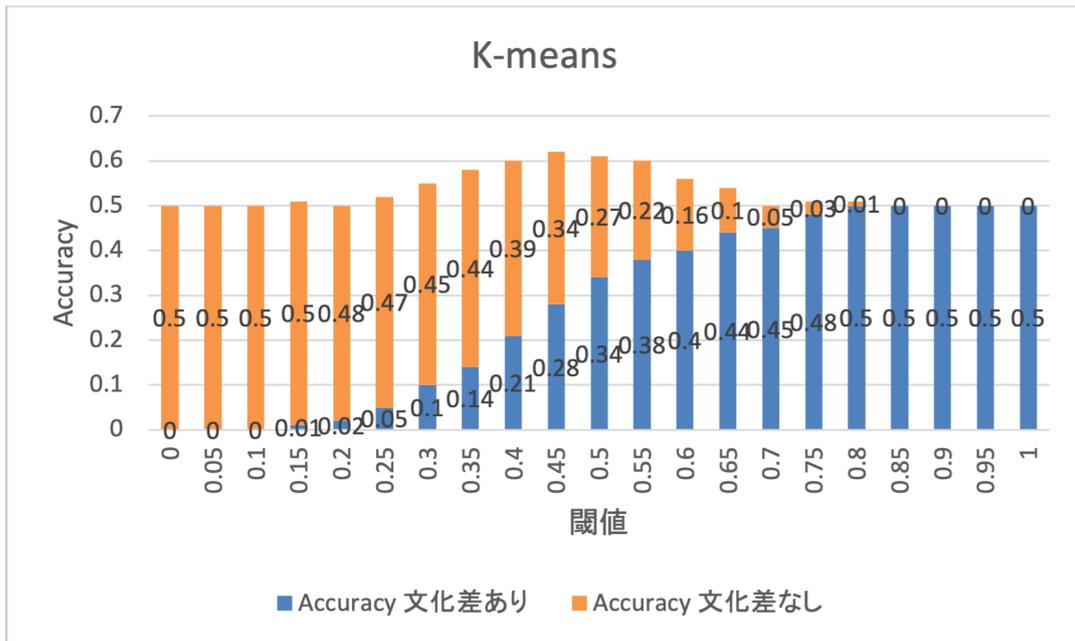


図 17: K-means 手法結果 (日インドネシア)

Accuracy が 0.8(80 %) であった。

これらの評価結果より、日インドネシア間でも従来の単純平均により平均特徴ベクトルを生成する手法が最も Accuracy が高かった (表 16)。

次に、分類した上位 3 層目の概念ごとに従来手法と改善手法の精度差を見てみる。表 17 は各概念のデータに対して、それぞれの手法の Accuracy を示したものである。この中で、改善手法の Accuracy が従来手法の Accuracy を上回った “causal agent”, “thing”, “group”, “relation” の概念のデータに注目する。これらのデータに χ^2 乗検定を用いて、従来手法と改善手法に有意な差があるかを評価した結果が表 18 である。自由度は 3 で “causal agent”, “thing”, “group”, “relation” の概念に対して、それぞれの手法は有意水準 5 % より大きいため有意差がない。そのため、各手法間で精度の差はないと言える。

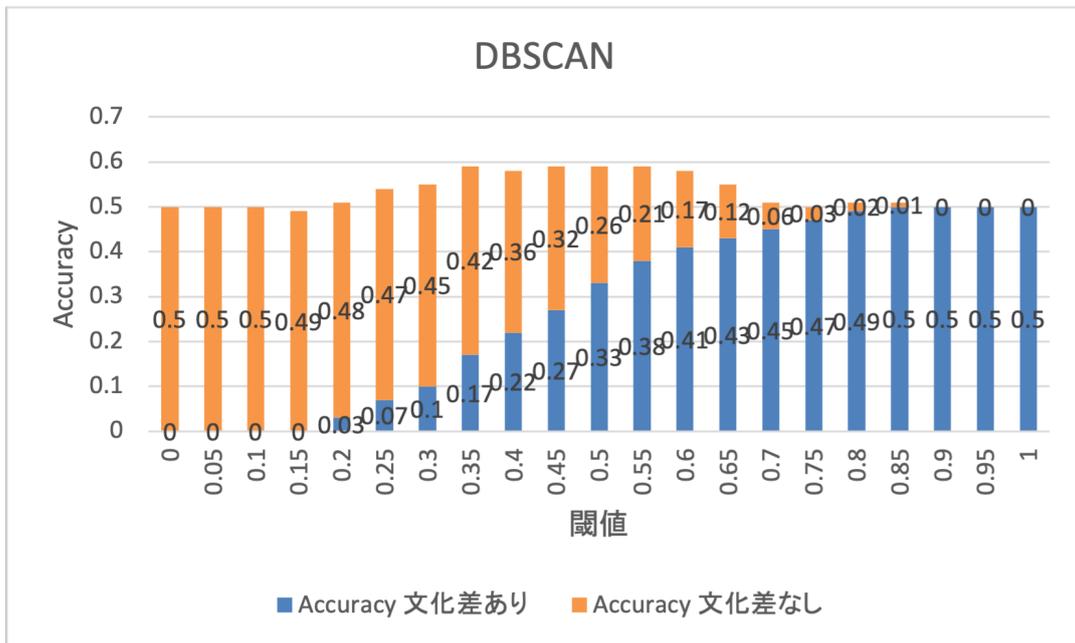


図 18: DBSCAN 手法結果 (日インドネシア)

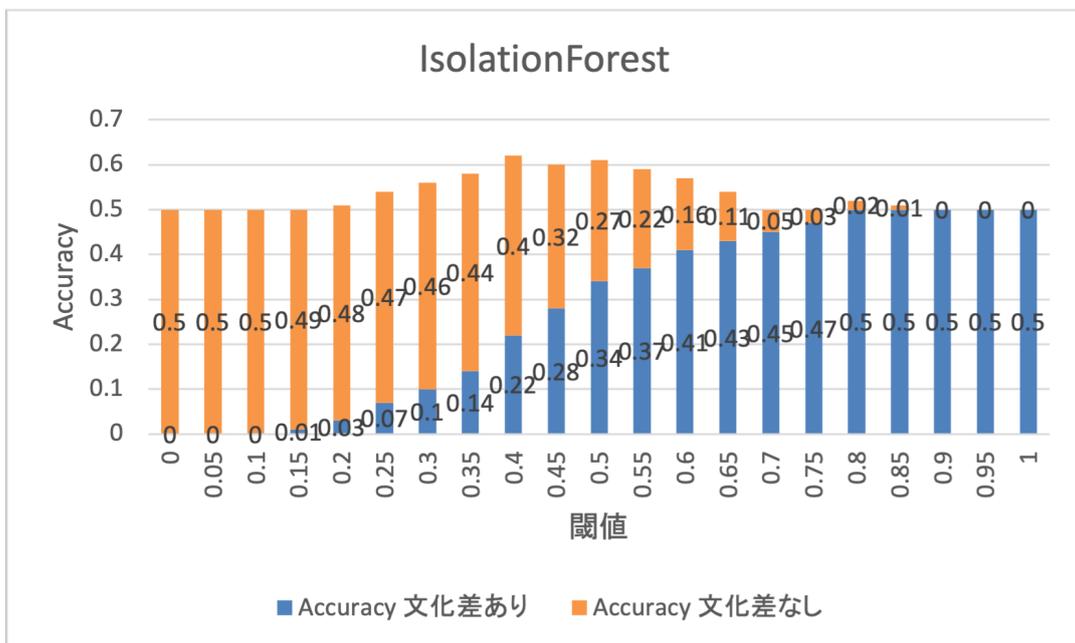


図 19: IsolationForest 手法結果 (日インドネシア)

表 17: 上位概念ごとに各手法での比較

概念	Accuracy(%)			
	従来手法	K-means	DBSCAN	IsolationForest
object	57	55	47	53
matter	63	61	57	63
process	70	70	60	70
causal agent	66	66	72	66
thing	59	65	63	59
psychological feature	62	52	54	56
attribute	71	57	64	71
group	67	67	65	69
communication	60	60	56	56
measure	58	58	53	58
relation	55	53	55	57

表 18: 各手法ごとの検定結果

層数	確率 (p)	χ^2 値
causal agent	0.883682282	0.655221376
thing	0.906027665	0.557772152
group	0.981552637	0.174789916
relation	0.98195314	0.172161172

第7章 考察

評価実験の日英間での文化差判定結果をもとに分析を行う。まず、作成した評価データセットと予備評価の時に使用したデータでどのような違いがあるかを見る。その次に、従来手法と改善手法の両方とも文化差検出を成功したデータと失敗したデータを述べる。その後、従来手法と改善手法で判定結果が異なったデータを説明し、従来手法と改善手法を比較して、今後の改善点を述べる。

7.1 評価データセットの分析

作成した評価データセットと予備評価で使用したデータの比較を行う。

予備評価では画像2つとキーワードを見てどちらの画像がよりキーワードから想起できるかを質問していた。そこから両方とも想起できるか、片方がより想起できるかという結果をもとに文化差ありと文化差なしを2値で判定していたため、画像単体での評価を得ることが難しかった。評価データセットでは各画像



図 20: ザクロ



図 21: pomegranate

ごとに説明とキーワードとの評価を得てからその対象概念の文化差を判定している。この違いによって、予備評価データと評価データセットで人手の文化差判定が異なった概念を挙げる。

まず、日本語で“ザクロ”，英語では“pomegranate, pomegranate tree, Punica granatum”という概念がある。この概念は予備評価のアンケート方法では“文化差あり”と人手で判定されている。これらの単語で取得された画像が図 20 と図 21 である。このように日本語のキーワードから取得された画像群ではザクロの果実の画像が多く、一方で、英語のキーワードで取得された画像ではザクロの木の画像が多く取得されている。この概念は予備評価のデータでは“文化差あり”データとなっているが、評価データセットでは“文化差なし”データと判定されている。予備評価データではその質問の方法から、画像に違いが見受けられた場合、キーワードからより想起できる画像が回答として選ばれることが考えられる。そのため、この“ザクロ”の例のようにどちらもザクロを指している画像ながら、果実の方がより想起されたため“文化差あり”判定となった。しかし、評価データセットでは各画像ごとに説明とキーワードから想起できるかを5段階で評価し、それらの評価の違いから文化差判定した結果、“文化差なし”判定に変わった。

次に、日本語で“ホルモン”，英語で“hormone”という概念がある(図 22 と図 23)。この概念は予備評価のアンケート方法では“文化差なし”データであったが、評価データセットでは“文化差あり”と判定されている。この概念では予備評価では日本語の画像も英語の画像も日本人には想起できると評価された。日本人には食用のホルモンの方が馴染みがあるが、それらが内臓を指す英語であるという情報も持っているため英語の画像も想起できると判定したと考えられる。しかし、評価データセットでは英語話者にとっては日本語の画像は想起できないため“文化差あり(包含関係)”という判定に変わった。

このように日本語話者だけでなく英語話者も含めて各画像の評価から文化差判定をした結果、より詳細で細かい文化差データを作成することができた。

7.2 文化差検出結果の例

次に、文化差検出に成功したケースと失敗したケースを紹介する。表 19 は実際に使ったデータの文化差判定結果の例である。

- 文化差なしデータを判定成功



図 22: ホルモン

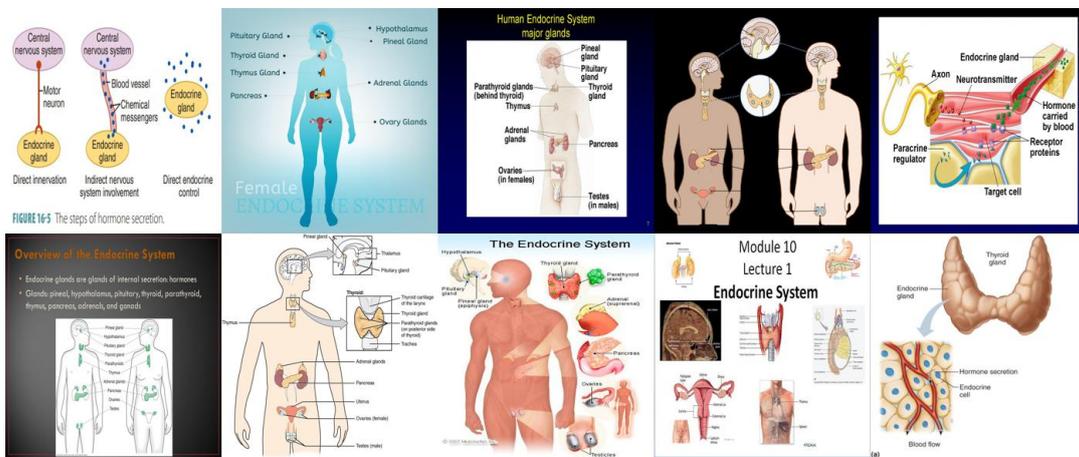


図 23: hormone

- 文化差なしデータを判定失敗
- 文化差あり (排他関係) データを判定成功
- 文化差あり (排他関係) データを判定失敗
- 文化差あり (包含関係) データを判定成功
- 文化差あり (包含関係) データを判定失敗

上からこのような順番になっている. 一番上の“ワイン”のケースでは, 文化差なしのデータを文化差なしと判定できたケースである. この時, 実際 to 取得された画像が図 24 と図 25 である. これはアンケートによって文化差なしと判定されたデータであり, 本研究の従来手法でも三つの改善手法でも全て文化差なしと判定できた例である. 具体的に算出された類似度が従来手法:0.702, K-means:0.593, DBSCAN:0.702, IsolationForest:0.702 であった. 実際 to 取得された画像を見てみると両方ともボトルにはいったワインの画像が取得されている. そのため, 類似度が高く算出された結果, 文化差なしと判定できたと考えられる.

表 19: 文化差判定の結果の例

判定結果	正解	キーワード		上位概念
		日本語	英語	
文化差なし	文化差なし	ワイン, ぶどう酒, ブドウ酒, 葡萄酒	wine, vino	causal agent
文化差あり	文化差なし	ザクロ	pomegranate, pomegranate tree, Punica granatum	object
文化差あり	文化差あり (排他)	フットボール, フットボール	football	object
文化差なし	文化差あり (排他)	プディング	pudding	matter
文化差あり	文化差あり (包含)	リンス	rinse	psychological feature
文化差なし	文化差あり (包含)	蠟涙, ワックス, 蝋	wax	things

次に、文化差なしデータを判定失敗したケースを紹介する。先ほど説明した“ザクロ”の例は判定に失敗したケースである(表 19)。実際に取得された画像群が図 20 と図 21 である。この時の類似度がそれぞれ従来手法:0.322, K-means:0.264, DBSCAN:0.311, IsolationForest:0.270 であった。取得された画像群では“ザクロ”で取得された画像がザクロの実の画像が多数である。一方で、“pomegranate, pomegranate tree, Punica granatum”で取得された画像はザクロの木であることが多数であった。英語ではキーワードの中に、“pomegranate tree”とあるため、ザクロの木の画像が多くなったと考えられる。これにより類似度が低くなり、文化差ありと判定されたと考えられる。



図 24: ワイン



図 25: wine

文化差あり (排他関係) のデータを正確に判定できたケースが表 19 の“フットボール”である。日本語から取得された画像群ではサッカーボールが多く取得されている (図 26)。一方、英語から取得された画像ではアメリカンフットボールの画像がより多く取得されている (図 27)。算出された類似度が従来手法:0.502, K-means:0.482, DBSCAN:0.476, IsolationForest が 0.473 であった。日本ではフットボールからはサッカーを連想し、アメリカンフットボールは“アメフト”と呼ばれることが多い。一方で、アメリカでは football からはアメリカンフットボールを連想する。そのため、画像に違いが生じ、それらのボールの形状の違いから類似度が低くなったため、文化差ありと判定できたと考えられる。



図 26: フットボール



図 27: football

反対に文化差あり (排他関係) のデータを判定失敗したケースが表 19 の “プディング” である. “プディング” から取得された画像では, 日本でいうマフィンが多数取得されている (図 28). “pudding” から取得された画像は日本でいうプリンやチョコプリンが多く取得されている (図 29). この時の類似度が従来手法:0.637, K-means:0.576, DBSCAN:0.637, IsolationForest:0.637 であった. 日本でいう “プリン” は和製英語でもともと英語の “pudding” が変換されてプリンになったと言われている. そのため, “pudding” は日本でいうプリンであるため, “pudding” からはプリンの画像が取得されている. 一方, 日本の “プディング” は米・肉・小麦粉などに牛乳や卵などを混ぜ, 煮たり蒸したりして固めた柔らかい食べ物の総称とされている [18]. そのため, “プディング” からはこのような画像が取得された. 本手法は画像から取得できる特徴をもとにベクトルを生成し類似度を算出している. そのため, 素材などは考慮しづらく, 同じような形をしているこれらの画像群の類似度は高く算出されたため, 文化差なしと間違った判定になったと考えられる.

次に文化差あり (包含関係) のなかで判定に成功したケースを紹介する. 表 19



図 28: プディング



図 29: pudding

の“リンス”では日本のリンスの画像が取得され, “rinse”からは水で洗っている画像が多数取得されている(図30・図31)。“rinse”の単語の意味は日本語でいう“すすぐ”という意味になり, 日本のリンスは英語では“conditioner”と言われる。実際にアンケート内の説明文では“石鹼を使わずに軽く洗うこと”となっている。そのため, 英語から取得された画像群と説明文が一致しているため, 日本人にとっては英語の画像群は想起できる。そして, “リンス”というキーワードによって日本語の画像群は想起できるため, 日本人は両方想起できると判定している。一方で, アメリカ人にとっては, 説明文もキーワードからも日本のリンスは想起できないため, 英語の画像群のみしか想起できない判定している。これらの結果から文化差あり(包含関係)になったと考えられる。また, この概念の類似度は従来手法:0.255, K-means:0.252, DBSCAN:0.255, IsolationForest:0.253であった。



図 30: リンス



図 31: rinse

次に文化差あり (包含関係) のなかで判定に失敗したケースを紹介する. 表 19 の“ワックス”では, 日本語から取得された画像には, ヘアワックスや革製品に使うワックスといったより商品的な画像が取得されている (図 32). 一方で, “wax”からは体に塗ったり, 引き伸ばされた蠟状の物質の画像が多数取得されている (図 33). そのため, 日本人から見ればどちらも想起できるが, アメリカ人から見ると日本の製品はわからず缶や箱にしか見れないため, 日本の画像は想起できなかったと考えられる. また, この時の類似度は従来手法:0.595, K-means:0.590, DBSCAN:0.595, IsolationForest:0.586 であった. これらは箱や缶など入れ物の形などが特徴としてベクトルに現れ, その影響から類似度が高くなったと考え



図 32: ワックス



図 33: wax

られる。

7.3 従来手法と改善手法の比較の例

従来手法と改善手法で判定が異なったケースを紹介する。まず、最初に紹介するのが日本語のキーワードが“群れ、一群れ、群鳥、群がり、一群、群”，英語の



図 34: 群れ



図 35: flock

キーワードが“flock”という文化差なしデータである。それらのキーワードから取得された画像が図 34 と図 35 である。どちらも羊や鳥などの群れの画像が多数取得されている中、日本語から取得された画像にも英語から取得された画像にもそれぞれのキーワードに引っ張られ、単語そのものが入った画像が取得されている。この時、従来手法では類似度:0.468 で文化差ありと文化差判定に失敗している。一方、異常値検出を用いた改善手法ではそれぞれ類似度が K-means:0.554, DBSCAN:0.622, IsolationForest:0.571 で文化差なしと正確に文化差判定することができている。

次に、日本語のキーワードが“処置, 手当, トリートメント, 手当て”, 英語のキーワードが“treatment, intervention”という文化差あり (包含関係) のデータである。それらのキーワードから取得された画像群が図 36 と図 37 である。アンケート結果では日本人にとっては両方の画像が想起できるが、アメリカ人にとっては“トリートメント”から取得された画像は想起できないであった。日本語から取得された画像では、言葉や花などが写った画像の中に、髪の毛に使用するトリートメントの画像が混在している。一方で、英語から取得された画像群で



図 36: トリートメント

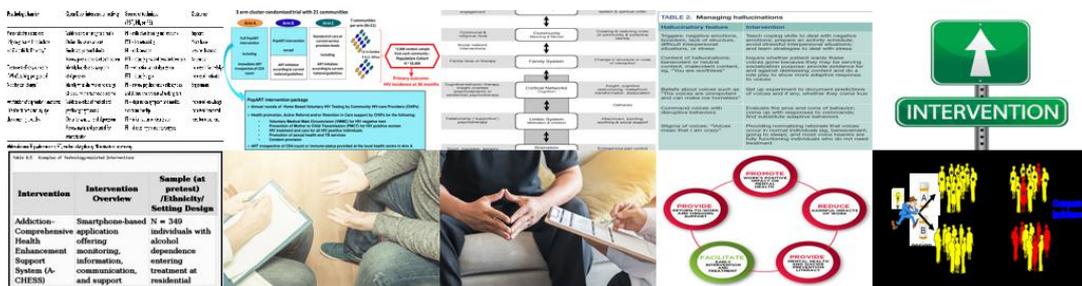


図 37: treatment

はカウンセリングの様子やキーワード自身が写った画像が取得されている。このデータでは従来手法は文化差判定に成功しているが、改善手法では全て文化差判定に失敗している。この時の Accuracy が従来手法:0.541, K-means:0.596, DBSCAN:0.664, IsolationForest:0.581 であった。このことから、日本で髪の毛に使用するトリートメントが異常値として検出されそれらの特徴がなくなったため、日英の両画像とも言葉が入った画像の特徴から類似度が算出されたため、類似度が高くなり、従来手法に比べて判定が悪くなったことが考えられる。

7.4 今後の改善

異常値検出を用いて、ターゲットと異なる画像を検出する手法を提案した。これにより 7.3 章の“群れ”のように画像検索の結果、キーワード自体が画像に入っている画像などノイズとなる画像を取り除くことによって、従来手法では文化差判定に失敗していたデータを正確に文化差判定することに判定することができた。一方で、全体的に Accuracy が従来手法よりも下がっていることから、判定結果が好転したケースよりも多く文化差判定結果が悪くなったケースが多い

ということがわかる。その原因として、そもそも画像検索結果が悪くない場合もクラスタリング等を用いて重みをつけていることが考えられる。現状の手法では、異常値検出で異常値となった画像は全て重みを低くしている。しかし、それらの判断基準は画像の特徴ベクトルがそれらの画像内で他のベクトルと比べて異常値であるかどうかで判断している。そのため、実際には文化差判定のために必要な情報を持った画像でもベクトルが離れているため、重みが低くなっている可能性がある。よって、今後の改善として異常値検出で検出されたデータが実際にノイズデータであるかを判定することが必要となってくる。

また、今回のアンケートでは文化差ありの中でも、包含関係か排他関係かを取得できる。しかし、現状の手法では文化差の有無しか文化差判定できないため、今後は文化差ありの中でもどのケースかを判定できるように改善する。そのために、画像だけではなく、言葉をもとにした手法を統合する。具体的には、説明文やキーワードを単語の分散表現を用いてベクトル化したものと画像から取得した特徴ベクトルをもとに統合ベクトルを生成し、各言語の統合ベクトル間で類似度を算出する。これによって、画像だけではわからない用途の違いなどを言葉から取得し、それらの情報を用いてそれぞれの文化差の関係を明らかにする。

また、“ザクロ”の例のように、キーワードが複数ある場合にそのうちの一つだけに該当する画像しか取得できていない可能性がある。現状の手法では、AND検索を用いることで単語が複数ある場合でもその概念を表す画像を取得しようとしている。しかし、各単語ごとに画像検索を行い、その単語ごとに画像から特徴ベクトルを生成し、最後に全ての特徴ベクトルを合成することで、これらの問題を解決できると考えられる。よって、今後はさらに対象概念を表す合成ベクトルの作成方法をさらに改善する。

第8章 おわりに

多言語コミュニケーションにおける文化差を解消するために、私たちは画像の特徴ベクトル間の類似度に基づく手法を提案してきた。その問題点として、ノイズとなる画像が含まれる可能性があるため、異常値検出を用いてノイズデータに影響を受けない文化差検出手法を提案してきた。本研究の貢献は以下の通りである。

文化差判定用の評価データセットの構築

WordNet の概念階層に基づき概念階層の上位3層目の概念がそれぞれ均等になるように概念を選択した。日英の両言語話者の視点から言葉と画像の複数の要素を評価をもとに差が見られる概念を文化差あり、差が見られない概念を文化差なしとしてラベル付けした。日英間で文化差ありが257件、文化差なしが565件獲得できた。また、1000個の概念を概念辞書の上位概念に基づいて分類した結果、分類した上位3層目の概念ごとに獲得できた文化差ありの個数は、object:16, matter:9, process:27, causal agent:22, thing:25, psychological feature:31, attribute:10, group:23, communication:40, measure:23, relation:31であった。

ノイズデータに頑健な文化差検出手法

作成した評価データセットを用いて提案手法を評価した結果、ノイズデータの除去や画像の確信度を用いない手法と比較して、Accuracyは向上しなかった。しかし、異常値検出を用いることで従来手法では文化差判定に失敗したデータを正確に文化差判定できるケースも見つかった。

日英インドネシアの言語話者の視点から言葉と画像を評価し、言語間で差が見られる概念を文化差あり、差が見られない概念を文化差なしとしてラベル付けした。これにより、日英間で文化差ありデータ257件、文化差なしデータ:565件を含んだ評価データセットを作成した。これを用いた評価実験では6割程度の確率で提案手法が正確に文化差判定できることが確認できた。また、画像検索により取得される画像にターゲットと異なる画像が含まれた場合、正確に文化差判定できない可能性があるため、異常値検出を用いてノイズとなる画像に頑健な文化差検出手法を提案した。従来手法に比べて、改善手法ではAccuracyが低くなったものの従来手法では文化差判定に失敗していたデータを改善手法では正確に文化差判定できるケースが見受けられた。しかし、異常値検出により重み付

けを行うことで文化差判定に必要な特徴を持った情報まで削ぎ落としてしまっている可能性が考えられる。そのため、今後の課題として、異常値検出により検出されたデータが実際に異常値なのかを判断し、異常値であった場合に重み付けすることが必要と考えられる。

謝辞

本研究を行うにあたり、ご指導いただきました村上陽平准教授, Mondheera Pituxcoosuvarn 先生に深く感謝申し上げます。また、アンケートにご協力いただきました皆様に感謝申し上げます。

参考文献

- [1] Pituxcoosuvann, M., Ishida, T., Yamashita, N., Takasaki, T. and Mori, Y.: Machine translation usage in a children’s workshop, *Proceedings of the International Conference on Collaboration Technologies*, Springer, pp. 59–73 (2018).
- [2] Pituxcoosuvann, M., Murakami, Y., Lin, D. and Ishida, T.: Effect of Cultural Misunderstanding Warning in MT-Mediated Communication, *Proceedings of the International Conference on Collaboration Technologies and Social Computing*, Springer, pp. 112–127 (2020).
- [3] Deutscher, G.: *Through the language glass: Why the world looks different in other languages*, Metropolitan books (2010).
- [4] 吉野孝, 宮部真衣ほか: 文化差データの収集サービスの提案, 第 75 回全国大会講演論文集, Vol. 2013, No. 1, pp. 515–516 (2013).
- [5] Yoshino, T., Miyabe, M. and Suwa, T.: A proposed cultural difference detection method using data from Japanese and Chinese Wikipedia, *Proceeding of 2015 International Conference on Culture and Computing (Culture Computing)*, IEEE, pp. 159–166 (2015).
- [6] 諏訪智大, 宮部真衣, 吉野孝ほか: 異文化間コミュニケーションにおける重要度を考慮した文化差検出手法の提案, 2014 年度 情報処理学会関西支部支部大会 講演論文集, Vol. 2014 (2014).
- [7] Pfeil, U., Zaphiris, P. and Ang, C. S.: Cultural differences in collaborative authoring of Wikipedia, *Journal of Computer-Mediated Communication*, Vol. 12, No. 1, pp. 88–113 (2006).
- [8] Hofstede, G. H., Hofstede, G. J. and Minkov, M.: *Cultures and organizations: Software of the mind*, Vol. 2, Mcgraw-hill New York (2005).
- [9] 喜多香織, 高崎俊之, 林冬恵, 中島悠, 石田亨ほか: 多言語知識コミュニケーションの分析と改善, 第 75 回全国大会講演論文集, Vol. 2013, No. 1, pp. 169–170 (2013).
- [10] Yanai, K., Yaegashi, K. and Qiu, B.: Detecting cultural differences using consumer-generated geotagged photos, *Proceedings of the 2nd International Workshop on Location and the Web*, pp. 1–4 (2009).

- [11] Heeryon, C., 石田亨, 山下直美, 稲葉利江子, 高崎俊之, 神田智子: 絵文字解釈における人間の文化差判定, *ヒューマンインタフェース学会論文誌*, Vol. 10, No. 4, pp. 427–434 (2008).
- [12] Cho, H., Ishida, T., Yamashita, N., Inaba, R., Mori, Y. and Koda, T.: Culturally-situated pictogram retrieval, *International Collaboration*, Springer, pp. 221–235 (2007).
- [13] Koda, T.: Cross-cultural comparison of interpretation of avatars’ facial expressions, *Proceedings of the IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)* (2006).
- [14] Isahara, H., Bond, F., Uchimoto, K., Utiyama, M. and Kanzaki, K.: Enhancing the japanese wordnet, *Proceedings of The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP*, Citeseer (2009).
- [15] Fellbaum, C.: WordNet, *The encyclopedia of applied linguistics* (2012).
- [16] Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [17] Pituxcoosuvann, M., Lin, D. and Ishida, T.: A Method for Automated Detection of Cultural Difference Based on Image Similarity, *Proceedings of the International Conference on Collaboration and Technology*, Springer, pp. 129–143 (2019).
- [18] 明松村: 大辞林 第四版, 三省堂 (2019).
- [19] Levinson, M. H.: Through the Language Glass: Why the World Looks Different in Other Languages, *ETC.: A Review of General Semantics*, Vol. 67, No. 4, pp. 473–474 (2010).
- [20] 諏訪智大, 宮部真衣, 吉野孝ほか: 日本語版・中国語版 Wikipedia を用いた文化差検出手法の提案, *情報処理学会論文誌*, Vol. 55, No. 1, pp. 257–266 (2014).
- [21] 吉野孝, 宮部真衣: 日本語版 Wikipedia における執筆者の意図に着目した日中間の文化差検出手法の検討, *電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション*, Vol. 111, No. 427, pp. 13–18 (2012).
- [22] Cho, H., Ishida, T., Yamashita, N., Inaba, R., Takasaki, T. and Koda, T.: 絵文字解釈における人間の文化差判定 Cho Heeryon* 1 石田亨* 1 山下直美稲葉利江子高崎俊之** 神田智子.

- [23] Cho, H., Ishida, T., Yamashita, N., Inaba, R., Takasaki, T. and Koda, T.: 絵文字解釈における人間の文化差判定 Cho Heeryon* 1 石田亨* 1 山下直美 稱葉利江子 高崎俊之** 神田智子.
- [24] Koda, T.: Cross-cultural study of avatars ' facial expressions and design considerations within Asian countries, *International Workshop on Intercultural Collaboration*, Springer, pp. 207–220 (2007).
- [25] Yamashita, N. and Ishida, T.: Effects of machine translation on collaborative work, *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pp. 515–524 (2006).
- [26] Pituxcoosuvarn, M. and Ishida, T.: Multilingual communication via best-balanced machine translation, *New Generation Computing*, Vol. 36, No. 4, pp. 349–364 (2018).

8.1 付録

アンケート1~25

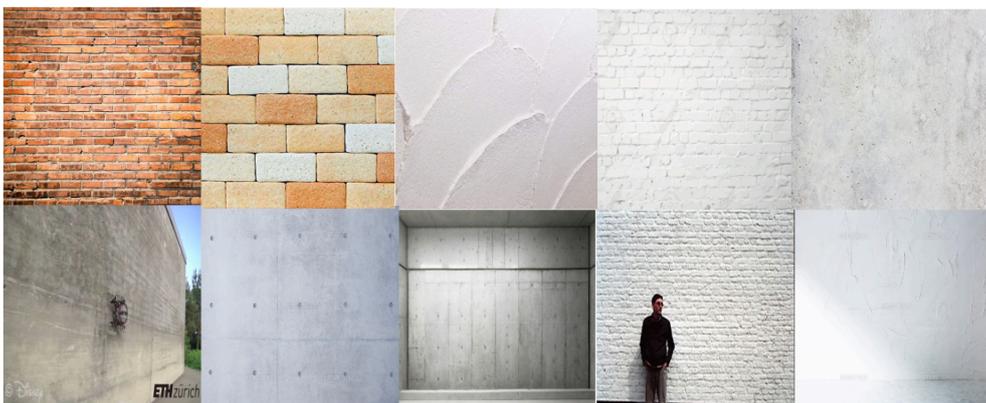
今から25組の画像群のセットを見てもらいます。

キーワードを見て、"A"と"B"のどちらの画像群がより連想できるか選んでください。

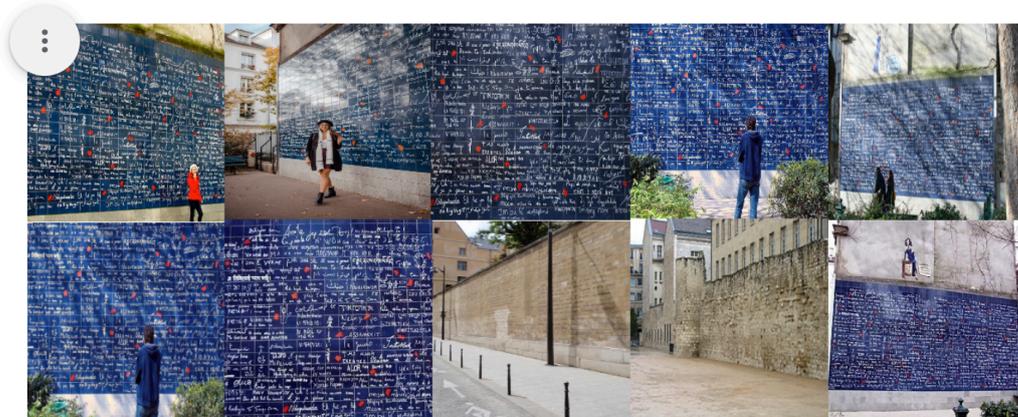
もし、両方とも連想できた場合は"両方"を選び、どちらも連想できない場合は"該当なし"を選んでください。

※()内の数字は質問の番号です

A (1) *



B (1)



キーワード：壁

- A 🖼️ ×
- B ×
- 両方 ×
- 該当なし ×

図 38: 人手による文化差判定用インタフェース

	データ1	データ2	データ3	データ4	データ5
1回目	V	V	V	V	T
2回目	V	V	V	T	V
3回目	V	V	T	V	V
4回目	V	T	V	V	V
5回目	T	V	V	V	V

V バリデーションデータ
 T テストデータ

図 39: 交差検証